Molecular characterisation of the bacterial community in dentinal caries

Schulze-Schweifing, Kathrin

Awarding institution:
King's College London

Molecular characterisation of the bacterial community in dentinal caries

Kathrin Schulze-Schweifing

A thesis submitted in accordance with the requirements of the degree of Doctor of Philosophy for King's College London

August 2012

Department of Clinical and Diagnostic Sciences, Dental Institute

King's College London

Abstract

Dental caries remains the major cause of loss of tooth function and associated morbidity. The development of new prevention and treatment measures requires a better understanding of the disease process, particularly the role of the associated microbiota. The aim of this study was to perform a comprehensive characterisation of the microbiota in dental caries using culture-independent and next generation sequencing methods, with a focus on improving the detection of bacteria with DNA of high G+C content. Oligonucleotide probes for taxa of interest were designed and evaluated using fluorescent *in situ* hybridisation (FISH). DNA extraction methods, PCR polymerases and primer pairs were compared for their ability to isolate and amplify 16S rRNA genes from species with low and high G+C content. DNA extraction method and polymerase were not found to significantly influence the detection of high G+C species but primer sequence was found to be of critical importance. Five "universal" PCR primer sets and culture were used to characterise the dentine caries microbiota of six subjects. From 3240 clones and isolates, 228 taxa representing eight phyla were identified by Sanger sequencing. Detection frequency of the high G+C *Actinobacteria* was 33% using culture, but 2.6-11.1% in the different clone libraries. The samples were analysed further by 454 pyrosequencing and 25758 sequences were identified to 264 taxa representing 11 phyla. Pyrosequencing allowed an analysis of greater depth, although the composition of the samples was comparable to those obtained by the Sanger-based method. Total and specific bacteria were detected successfully in excavated dentine by FISH with universal and specific probes, while in bisected teeth only total bacteria were seen with the universal probe. In conclusion, this study has revealed a highly diverse caries microbiota. Pyrosequencing increased detection of taxa and overall coverage, but all methods had associated biases which affected the results obtained.

Acknowledgements

I share the credit of my work with many people who have helped, encouraged and supported me in a myriad of ways.

I would like to express my sincere gratitude to my supervisors Prof William Wade and Prof Avijit Banerjee, who have encouraged and criticised as and when needed and helped me navigate the sometimes cloudy, sometimes stormy, waters that is a research thesis.

My deepest gratitude to all patients who have kindly provided the samples on which this research is based. A very big thank you also goes to the lovely Mona Hajatdoost-Sanj, who helped recruit volunteers and co-ordinated collection of all patient samples. In addition, Pete Pilecki gave much appreciated practical support in bisecting the collected teeth and Ahmed Almahdy kindly helped excavate caries scoops.

Fellow PhD student James Kistler organised and carried out the experimental work needed for the 454 pyrosequencing in collaboration with the Centre for Haemato-Oncology, Barts Cancer Institute, Queen Mary University of London. Many thanks not just for the work, but also the numerous questions you've patiently answered.

I would also like to thank Dr Dave Moyes for descending on my writing with a sharpened red pen, giving very helpful and much appreciated constructive criticism.

Many thanks also go to my friends and colleagues (present and past), not just of the Micro Department, but also the Salivary Research and Mucosal Immunity, Fungal and HIV infection groups – thank you guys for your friendship, many pieces of advice and generally for making the labs a fun place to work.

I am also grateful to be blessed with a fabulous (extended) family and many friends outside work who have boosted my morale on various occasions over the last few years.

Finally, I would like to dedicate this thesis to my wonderful husband, without whom this thesis would simply have remained a dream. Christian, your encouragement, support and understanding has made me believe in myself and enabled me to chase my dream. The love and gratitude I feel for you and our precious children cannot be expressed in mere words.

Table of Contents

List of Figures

List of Tables

List of Abbreviations

ATP	adenosine triphosphate
ABC	ATP-binding cassette
BLAST	basic local alignment search tool
BA	blood agar
bp	base pair
BHI	brain heart infusion broth
CARD-FISH	catalysed reporter deposition fish
CCD	charge coupled device
CFB-cluster	Cytophaga, Flavobacterium and Bacteroides cluster
CLSM	confocal laserscan microscopy
cfu	colony forming units
CMV	cytomegalovirus
dNTPs	deoxynucleotides
ddNTPs	dideoxynucleotides
DLVO model	named after Derjaguin, Landau, Verwey and Overbeck
DMSO	Dimethyl sulfoxide
DNA	deoxyribonucleic acid
ECC	early childhood caries
EDJ	enamel-dentine junction
EBV	Epstein-Barr virus
EtOH	ethanol

EPS	exopolysaccharides
FISH	fluorescence in situ hybridisation
HSV	erpes simplex virus
HOMD	human oral microbiome database
HERVs	human endogenous retroviruses
HOMIM	human oral microbe identification microarray
HPV	human papilloma virus
HA	hydroxyapatite
IL	interleukin
Km	kanamycin
LB	Luria Bertani medium
LNA	locked nucleic acids
MID	multiplex-identifier sequence
ml	millilitre
mICDAS	modified International Caries Detection and Assessment System
MG1	mucin glycoprotein 1
MG2	mucin glycoprotein 2
MS	mutans streptococci
NCPs	non-collagenous proteins
non-MS	non-mutans streptococci
nseqs	number of sequences
nt	nucleotide

OD	optical density
OTU	operational taxonomic units
PCMX	para-chloro-meta-xylenol
PBS	phosphate buffered saline
Pfu	*Pyrococcus fumarosis*
PTP	picotiter plate
PCR	polymerase chain reaction
PNAG	poly-N-acetylglucosamine
PIA	polysaccharide intercellular adhesin
PRPs	proline-rich proteins
PPi	pyrophosphate
RDP	Ribosomal Database Project
RNA	ribonucleic acid
RTM	reduced transport medium
sIgA	secretory Immunoglobulin A
Sobs	observed species
Taq	*Thermus aquaticus*
TMAC	tetramethylammonium chloride

Chapter 1

Introduction

1 Chapter 1

Historically, the first reference to dental decay was found on a clay tablet excavated in the Euphrates Valley dating from 5000 BC, where 'the legend of the toothworm' was described (Suddick et al. 1990). In the early 17th century Antonie van Leeuwenhoek examined supragingival plaque from his own teeth under a microscope and revealed 'animalcules', now known to be oral bacteria (Jenkinson et al. 2005). It has since been established that the human oral cavity harbours a plethora of microorganisms, but much about dental caries initiation, progression and significant associated interspecies interactions remains unknown.

1.1 Microbiota of the human oral cavity in health

The oral cavity includes a variety of different surfaces, such as the teeth, tongue, gingivae and buccal mucosa, all of which are colonised by a variety of microorganisms including bacteria, archaea, fungi, protozoa and viruses. Many of these organisms exist as commensals in the oral cavity; however, for many microbes an association with disease has been established following Koch's work on causes of disease and the subsequent publication of his postulates.

Bacteria and *Archaea* are microorganisms with a circular chromosome and most have no nuclear membrane, while the eukaryotic microorganisms such as fungi and protozoa have linear chromosomes that are contained in a nuclear membrane. *Archaea* are widespread in nature and have been detected in the oral cavity, although only in very low numbers compared to bacteria. Research targeting archaea has intensified in recent years, as it was shown that some organisms from this domain, primarily members of the genus *Methanobrevibacter*, were associated with periodontal disease (Horz et al. 2011; Matarazzo et al. 2011). A recent study, however, detected *Archaea* in healthy volunteers as well as patients suffering from generalised aggressive periodontitis (Matarazzo et al. 2011).

The carriage rate of fungi, primarily members of the genus *Candida*, in healthy/asymptomatic individuals is around 35 %, but can be up to 55 % in hospitalised patients, or patients with oral prostheses (Bagg et al. 1999). *Candida albicans* is the main causative agent of fungal infections in the oral cavity. Infection generally only occurs when the microbiota becomes imbalanced due to, for example, impaired

salivary gland function, diabetes mellitus or immunosuppressive conditions such as HIV/AIDS (Akpan et al. 2002). In health, *C. albicans* is restricted to growth on the surface of mucosal cells, but in the case of infection the yeasts produce hyphae and invade the host cell via 'induced endocytosis' the end result of which is tissue damage via necrosis and/or apoptosis (Naglik et al. 2011).

Two species of protozoa are found in the oral cavity: *Entamoeba gingivalis* and *Trichomonas tenax*. They can be found in healthy mouths, but are found more frequently in individuals with periodontal disease. They are primarily saprophytic, feeding on food debris and bacteria in plaque, and are not thought to play any causal role in disease (Beatman 1933; Ghabanchi et al. 2010; Onyido et al. 2011).

Potentially pathogenic viruses, such as human herpes virus (HHV) type 1, 6, 7 and 8, Epstein-Barr virus (EBV) and cytomegalovirus (CMV), can be detected in the saliva of healthy individuals (Di Luca et al. 1995; Miller et al. 2006). *Herpes simplex* is one of the most important viruses in the oral cavity with a carriage rate of about 50 % by age 30, and causes herpes labialis, or cold sores. Herpes Simplex Viruses (HSV) have also been associated with carcinogenesis, since antibody levels to HSV-1 and -2 were higher in oral cancer patients when compared with controls and HSV seropositivity together with smoking has been associated with increased cancer risk (Meurman 2010). Human papilloma virus (HPV) has also been associated with oral cancer (de Villiers et al. 1985). Studies reported positivity rates for this virus of between 10 and 20 % in oral cancers, while in tonsillar carcinomas HPVs were detected in over 50 % of tested biopsies (zur Hausen 1996). In the case of the human immunodeficiency virus (HIV), which can be detected in saliva of infected subjects, the oral cavity offers protection against transmission through endogenous salivary inhibitors of HIV on the one hand; on the other hand oral lesions can be indicative of HIV infection in previously undiagnosed cases and predict disease progression in known infected subjects (Chapple et al. 2000).

Traditionally, investigations have tried to determine the bacterial causative agents of diseases by cultivating them with artificial media, but not all bacteria can be cultured in the laboratory, as will be discussed below. Therefore, efforts have been made in recent years using state-of-the-art molecular techniques to improve our knowledge of

the commensal microbiota in health enabling us to define processes and species contributing to disease initiation and progression.

Aas et al. (2005), using 16S rRNA gene amplification, cloning and sequencing, examined nine different sites in the oral cavities of five clinically healthy subjects and found that anatomic differences were the primary determinant on local microbiota composition. Only a few species were commonly found at most sites (for example *Streptococcus mitis* which was detectable at all sites of all subjects), while others appeared to be site-specific (for example *Rothia dentocariosa*, *Actinomyces* spp. and *Streptococcus sanguinis* preferentially colonising teeth) or subject specific (*Prevotella*) (Aas et al. 2005). A similar study attempted to define the healthy core microbiome of oral microbial communities using high throughput sequencing and found that the major proportion of the microbiome in three healthy and unrelated patients was shared between them. The predominant genera found were *Streptococcus*, *Veillonella*, *Granulicatella*, *Neisseria*, *Haemophilus*, *Corynebacterium*, *Rothia*, *Actinomyces*, *Prevotella*, *Capnocytophaga*, *Porphyromonas* and *Fusobacterium* (Zaura et al. 2009). These findings were confirmed by Bik et al. (2010), who also detected these genera in their study looking at the bacterial diversity in the oral cavity of ten healthy individuals. Moreover, they also identified *Cardiobacterium*, *Atopobium*, *Bergyella* and members of the phylum TM7 as part of the core oral microbiome in health. Further, since only eleven bacterial species were found in all of the study participants they concluded that each individual's microbiome is unique whilst the microbiota amongst different individuals showed more similarity at the genus level. This observation of great individual variation in colonisation was also found in many other studies (Listgarten et al. 1975; Munson et al. 2004; Nadkarni et al. 2004; Chhour et al. 2005). As a result, it has been suggested that the core microbiome might be better defined by looking at community function rather than community membership (Bik et al. 2010).

1.2 Oral diseases and their associated microbiota

Changes directly or indirectly affecting the environment of the health-associated microbiota can trigger a shift in population dynamics, as will be discussed below. Consequently, a range of diseases affecting the oral cavity is associated with the normal or commensal oral microbiota, including dental caries, gingivitis, periodontitis, and fungal infections. None of these diseases are infections in the classical sense, in

that they are not caused by exogenous organisms that enter the body and cause disease, but are of endogenous origin. This phenomenon is also referred to as amphibiosis, and an amphibiont or a symbiont undergoing dynamic change holds a critical position between mutualistic symbiosis and pathogenicity (parasitic symbiosis)(Ruby et al. 2007). This phenomenon is by no means exclusive to the oral cavity, and another well-known example includes the gut microbiota. Hippocrates has been quoted as saying, 'death sits in the bowels' and 'bad digestion is the root of all evil' in 400 B.C., showing that the importance of the intestines (and its microbiota) in human health has been long recognised and many studies have focused on this area (Sekirov et al. 2010). Further examples of endogenous diseases, amongst others, include disorders of the skin (Grice et al. 2011) and human endogenous retroviruses (HERVs) (Ryan 2004).

The periodontal diseases include a range of pathologies affecting the supporting structures of the tooth, such as the gums and bones in which the tooth is anchored. The initial stage in these diseases is the development of gingivitis. Gingivitis is an inflammation of the gingivae and is essentially a non-specific condition in that no individual bacterial species are implicated. However, the severity of the inflammation is related to the amount of plaque present (Theilade et al. 1966; Theilade 1986). In addition, as plaque matures it becomes more anaerobic, resulting in an increase in the numbers of Gram-negative bacteria. The presence of large numbers of Gram-negative bacteria with their intrinsic endotoxin is subsequently thought to contribute to the inflammation (Schultz-Haudt et al. 1954; Tanner et al. 1998). Studies looking at experimental gingivitis revealed the presence of primarily Gram-positive cocci and rods in health, whereas three days after the suspension of oral hygiene Gram-negative cocci and rods appeared, as well as filaments and fusobacteria, leading to a further increase in Gram-negative bacteria and the appearance of spirochetes (Loe et al. 1965; Theilade et al. 1966).

If gingivitis is left untreated, periodontitis, which is irreversible, can develop in susceptible individuals. The disease develops from a combination of bacterial attack on the gums and the host's immune-inflammatory response to it. Periodontitis is characterised by loss of attachment between the gingivae and teeth with the formation of periodontal pockets as a reaction to the continued inflammation

(Listgarten 1986). Unlike gingivitis, certain species, including *Porphyromonas gingivalis*, *Tannerella forsythia* and *Treponema denticola,* the so-called red complex of periodontitis associated species described by Socransky *et al.* (1998), have been associated with disease severity. Other bacteria have also been associated with other forms of periodontitis. For example, *Aggregatibacter actinomycetemcomitans* is specifically associated with localised aggressive periodontitis in young adults (Zambon et al. 1983; Zambon 1985).

Periodontal abscesses can develop from established periodontal pockets, which can result in the destruction of soft tissue and bone in extreme cases (Bagg et al. 1999). Associated microbiota were found to be composed of periodontal pathogens such as *Fusobacterium spp.*, *P. intermedia/nigrescens, P. gingivalis* and *A. actinomycetemcomitans* (Jaramillo et al. 2005).

Finally, dental caries is the dissolution of tooth structure by acids formed by bacteria as a result of the fermentation of dietary carbohydrate, particularly sucrose. The pathogenesis of dental caries, the subject of this thesis, will be dealt with in detail in section 1.9.4 (Microbiology of Caries). Untreated caries can result in a range of complications, most commonly endodontic infection resulting from bacteria gaining access to the pulp chamber. When this occurs, the root canal is colonised by a diverse mix of anaerobic bacteria and their toxins cause inflammation and pus formation (Robertson et al. 2009). As with diseases such as periodontitis, no single causative pathogen has been identified (Munson et al. 2002; Robertson et al. 2009). Indeed, a study by Munson et al. (2002) investigating the microbiota in endodontic infections using molecular as well as culture methods found more taxa and a more diverse microbiota per sample than previously reported in studies using culture techniques alone.

As well as the gingivae and teeth, salivary glands can also become infected by either viruses (mumps, for example) or bacteria, such as *Staphylococcus aureus* and alpha haemolytic streptococci, resulting in parotitis or submandibular sialadenitis (Bagg et al. 1999).

1.3 Shifts in the oral microbiota as a result of dynamic interactions with the host and/or environment

Over time, several hypotheses explaining the aetiological role of plaque bacteria in caries and periodontal diseases have been proposed. The non-specific plaque hypothesis stated that the cause of disease was due to the overall interaction of plaque microbiota with the host (Theilade 1986). In contrast to this the specific plaque hypothesis stated that only very few species out of the resident plaque microbiota are actively involved in causing disease (Loesche 1976). However, it was argued that the specific hypothesis could not explain occurrence of disease in the absence of certain pathogens (e.g. caries lesions in which *S. mutans* could not be detected) or, conversely, absence of disease in the presence of known pathogens (Marsh 1994). It has been argued by Marsh (1994) that plaque-mediated diseases, while not necessarily having a totally specific aetiology, do show evidence of specificity and therefore proposed the ecological plaque hypothesis in an attempt to incorporate observations made in the laboratory and clinics with previous hypotheses. In this hypothesis a change in environmental factors is attributed to causing a shift in the balance of the prevailing plaque species, which in turn, could predispose a habitat to disease (Marsh 1994).

A number of factors can induce a shift from mainly health associated microbiota to mainly disease associated ones, resulting in periodontal disease or dental caries. One important factor is the maintenance of good oral hygiene. Measures of oral hygiene used by most people in developed countries are brushing of teeth using a toothbrush and toothpaste. Use of dental floss, which helps remove plaque and food debris from interdental spaces that are hard to reach using a toothbrush, is also recommended. Loe et al. (1965) performed a series of experimental gingivitis trials in humans, demonstrating that stopping oral hygiene results in the accumulation of plaque and initiates inflammation in all subjects. Reinstatement of oral hygiene and removal of accumulated plaque reverts the gingivae back to a status of health within a few days (Loe et al. 1965; Theilade et al. 1966).

Loe et al. (1965) observed variation in the length of time it took for gingivitis to develop in the various patients, which hints at the fact that different people exhibit diverse susceptibility to periodontal disease. In fact, it was found in segregation analyses of families with periodontal disease that inheritance mechanisms played a

role, but no single mode of inheritance that would include all affected families could be established (Meng et al. 2007). Many studies have examined genetic and immunological links to disease susceptibility and prevalence. Gene polymorphisms in interleukin-1, IL-4 and human leukocyte antigen, amongst others, have been examined as potential markers for increased susceptibility of aggressive periodontitis, but results were inconclusive (Lopez et al. 2005; Meng et al. 2007). Niederman et al. (2001) looked at the possible roles of a deficiency in phagocygote function, a pathogenic oral biofilm and/or deregulated gingival cytokine expression in the aetiology of early-onset periodontal disease. Their hypothesis was tested in P/E-selectin molecule deficient mice that mimic the human leukocyte adhesion deficiency and they found such mice to exhibit spontaneous early onset alveolar bone loss, a ten-fold elevation in bacterial colonisation of the oral cavities and elevated gingival tissue levels of the bone resorptive cytokine IL-1α (Niederman et al. 2001). Prophylactic administration of antibiotics was found to prevent alveolar bone loss. Dayan et al. (2004), on the other hand, found that transgenic mice that over-expressed the 17kDa form of IL-1α in the basal layer of oral mucosal epithelium develop all the clinical features of periodontal disease, while prophylactic treatment with antibiotics had no effect on disease severity in these animals.

Genetic links have also been examined with regards to the aetiology of dental caries. Bretz et al. (2005), looking at dental caries and microbial acid production in twins, concluded that variation in dental caries surface traits have a strong genetic component and that microbial acid production is modulated by the environment. A more recent study by Deeley et al. (2008) examined single-nucleotide polymorphism markers of genes that influence enamel formation in a study pool of unrelated humans and found that having at least one copy of the rare amelogenin marker allele was associated with increased age-adjusted caries experience.

As well as genetics, other environmental factors can also affect an individual's health. It is generally accepted that stress can affect immune functions and increase susceptibility to infectious diseases. Physical or mental stress can result in depressed immune responsiveness, which in turn influences the stress-related hormones glucocorticoid hormones. Glucocorticoids, which can also have an immunosuppressive function, are present in the gingival crevicular fluid and may contribute to periodontal

destruction due to being a source of nutrients for periodontal pathogens (Meng et al. 2007). Bosch et al. (2003) reported that even moderate stressors had the ability to alter the activity of mucosal secretory glands and could thereby affect microbial colonisation and resulting from that, susceptibility to disease.

Other lifestyle factors, such as diet and smoking, are further important factors influencing the composition of the oral microbiota. It is well established that the consumption of fermentable carbohydrates increases the likelihood of incidence of dental caries by the production of acid by bacteria (Parisotto et al. 2010; Huew et al. 2012). Smoking also has various effects on the host-immune system, such as a decreased IFN-gamma response, which probably aggravates the decreased IL-2 responses of patients with generalised aggressive periodontitis to *Porphyromonas gingivalis* and *Fusobacterium nucleatum* (Borch et al. 2009), or tobacco induced alterations to *P. gingivalis*-host interactions consisting of lower pro-inflammatory response (tumour necrosis factor-α, interleukin-6, interleukin-12 p40) from monocytes and peripheral blood mononuclear cells (Bagaitkar et al. 2009).

Systemic diseases can furthermore affect the oral microbiota. For instance, in Crohn's disease and ulcerative colitis, extra-intestinal involvement of, for example, the oral cavity, is well documented and can result in several types of lesions (Lourenco et al. 2010). Diabetes mellitus type 2 also has a strong association with periodontal disease; so much so, that it was called the 6th complication of diabetes mellitus (Loe 1993; Pihlstrom et al. 2005; Khader et al. 2006). Xerostomia (dry mouth) can have numerous aetiologic factors ranging from side effects of medications, smoking, alcohol abuse, to radiation therapy. Systemic diseases potentially causing dry mouth are Sjögren's syndrome, diabetes mellitus, HIV, Hepatitis C, Parkinson's disease, certain tumours and others (Ram et al. 2011). These conditions have been reported to result in increased development of dental caries due to an increase in *S. mutans* and *Actinomyces* species (Llory et al. 1972; Brown et al. 1975). Yeast counts have also been found to be raised and can result in fungal infection (Llory et al. 1972; Guobis et al. 2011).

1.4 Oral surfaces and microbial habitats

The vast majority of the oral microbiota reside on surfaces within the oral cavity of which there are three types of surfaces available for bacterial colonisation: the teeth, epithelial mucosa and bacterial/biofilm surfaces on the other two surfaces (Whittaker et al. 1996). The many habitats such as tooth enamel, tongue papillae, gingival crevices and mucosal membranes are each colonised by characteristic bacterial consortia (Aas et al. 2005).

Teeth consist of hard, non-shedding surfaces that provide distinct habitats for microorganisms, e. g. fissures, smooth surfaces, gingival crevices, etc. and make up about 20 % of the available oral cavity surface for microbial colonisation, facilitating the development of the biofilm known as dental plaque (Marsh et al. 1999; Marsh 2000; Socransky et al. 2005). The tooth enamel, made up of a mineral (95-96 % w/w) and an organic (1 % w/w) phase as well as water (3 % w/w), is the hardest substance in the human body to which only very few bacteria can directly attach (Orstavik et al. 1974; Frandsen et al. 1991; Mann et al. 2006).

The epithelial mucosa cover the lips, cheeks, palate and tongue, providing a multi-layer, stratified surface. These surfaces provide an ideal attachment site for microbes and throughout the body are the main point of contact between microbes and their host. For example, the papillary structure of the dorsum of the tongue provides refuge for many microorganisms which would otherwise be removed by mastication and the flow of saliva (Marsh et al. 1999).

1.4.1 Saliva, the acquired pellicle and the oral microbiota

Unlike skin, the epithelial mucosa of the oral cavity are kept moist continuously by secretions of saliva. Saliva is a fluid secreted by the salivary glands in the oral cavity. It is composed of more than 99 % water and contains a variety of electrolytes including sodium, potassium, calcium, magnesium, bicarbonate and phosphates, as well as immunoglobulins, proteins, enzymes and mucins (Humphrey et al. 2001). Saliva has manifold characteristics, including buffering properties (pH 6-7), lubrication as well as a role as a protective agent and plays a significant role in human metabolism and the development of disease or its cure (Scannapieco 1994; Marsh 2000; Humphrey et al. 2001). It is, in any case, a prerequisite for the maintenance of a healthy oral cavity

(Scannapieco 1994; Bosch et al. 2003), and several studies suggest that an altered salivary flow influences microbial composition and therefore the state of health or disease (Almstahl et al. 1999; Marsh 2000; Haffajee et al. 2005; Ruby et al. 2007).

Salivary proteins interact with the microbial cells in a manner that either shows antimicrobial characteristics through the removal of cells and even bactericidal effects or facilitates their colonisation through the provision of nutritional compounds and promotion of cell-surface attachment (Jong et al. 1987; Scannapieco 1994). Experiments have shown that although free amino acids are available as nutritional compounds in saliva, streptococci cannot utilise these and hydrolyse whole salivary proteins like lysozyme, lactoferrin, and lactoperoxidase (Dumas et al. 1987).

Salivary antimicrobial molecules such as histatins (MacKay et al. 1984), cystatins (Baron et al. 1999), lactoferrin (Aguirre et al. 1993) and salivary peroxidases (Gothefors et al. 1975) kill bacteria, but no connections could be drawn between the concentration of any of these components in saliva and oral disease conditions (Rudney et al. 1991). It has been found that saliva composition and therefore the efficiency of cell-agglutination and their clearance varies between individuals, since saliva from caries-resistant subjects was more effective in, for example, agglutination of *Streptococcus sanguinis* (Rosan et al. 1982).

A further important role of saliva and its macromolecules is the formation of the acquired pellicle. In 1839, Alexander Nasmyth described for the first time what he called ''the persistent dental capsule'', a membrane that detached from the tooth after acid treatment (Lendenmann et al. 2000). This membrane has subsequently been called the conditional film in environments other than the oral cavity and the acquired pellicle in the mouth (Lendenmann et al. 2000; Yao et al. 2001).

The formation of the acquired pellicle on a clean tooth surface is considered the first stage of dental plaque formation. The pellicle, a 0.1-0.7 µm thick layer of proteins, glycoproteins and other macromolecules, forms rapidly after cleaning of the teeth and is the result of non-random salivary protein adsorption to hydroxyapatite due to ionic interactions (Meckel 1965; Morge et al. 1989; Hannig et al. 2006; Siqueira et al. 2007). Forces such as electrostatic interactions between the salivary proteins' charged groups and surface ions like calcium and phosphate determine which proteins can attach.

Further factors in the pellicle development are site-specific variations in saliva composition and secretion, as well as van der Waals and hydrophobic interactions, since they result in a gain of entropy (Scannapieco 1994; Hannig et al. 2006). Statherin, histatin and proline-rich proteins are a few examples of proteins found to attach early in the process of pellicle development and belong to the pool of more than 100, mostly hydrophobic, proteins and peptides that are associated with this membrane (Lendenmann et al. 2000; Hannig et al. 2006; Siqueira et al. 2007). Although the pellicle primarily consists of saliva, its molecular composition differs from that of pure saliva in that it also contains gingival crevicular fluid, bacterial fragments as well as mucosa (Kolenbrander et al. 1993; Scannapieco 1994).

The pellicle plays an important role in bacterial attachment, since it contains molecules such as α-amylase and proline-rich peptides that act as receptors and promote adherence of specific species (Whittaker et al. 1996; Marcotte et al. 1998). Simultaneously, other species are prevented from adhering as they lack the appropriate receptors for the pellicle proteins (Scannapieco 1994; Marcotte et al. 1998; Bosch et al. 2003). The composition of the pellicle varies with secreted protein concentrations, which may influence the colonisation pattern and therefore the ratio of commensal to pathogenic species (Scannapieco 1994). However, not only complex molecules influence binding properties. Even single ions, like calcium, can influence bacterial attachment, since these ions can bridge surfaces of the same electrical charge and thus promote unspecific binding (Scheie 1994).

After proteins have been immobilised to hydroxyapatite, the primary mineral crystal of which enamel is comprised, they are subjected continuously to modifications by microbial or host enzymes that affect their chemical and enzymatic properties (Hannig et al. 2005; Hannig et al. 2006). Indeed, the endothermic process of binding to the enamel surface that is driven by an increase in entropy can trigger conformational changes in the proteins resulting in the exposure of hidden receptors that are also referred to as cryptitopes (Gibbons 1989; Hannig et al. 2005; Hannig et al. 2006). The exposure of cryptitopes is thought to be due to interactions of the negatively charged amino terminal sequence of the molecule with calcium ions on the surface that subsequently unfold the polypeptide chain (Gibbons 1989). However, the immobilisation of proteins is not permanent and pellicle formation and maturation is

characterised by the continued adsorption and desorption of molecules (Hannig et al. 2006; Siqueira et al. 2007).

1.5 Plaque formation

One of the first steps of plaque formation, alongside the development of the acquired pellicle, is the initial transport of bacterial cells to a surface, which can take place by Brownian motion, liquid flow, chemotaxis or motility. The subsequent attachment of bacteria to oral surfaces can be divided into two basic steps: The first stage of bacterial attachment is non-specific and reversible, involving only weak interactions between the microbial cells and pellicle components. The second stage of adhesion is mediated by bacterial ligands that specifically bind pellicle receptors. Once the primary colonisers are attached, the plaque develops by coaggregation and bridging of microorganisms in specific interrelations. Finally, the plaque composition and structure matures through bacterial growth, exopolymer production and further adhesion as well as detachment of cells.

1.5.1 Reversible adhesion

Several theories have been published regarding the physico-chemical properties of primary, or reversible, adhesion. The thermodynamic approach is based on Gibbs law of free energy, which states that adhesion takes place if energy is minimised in the process, resulting in a gain in entropy. This is described as:

$$\Delta_{ads}G = \Delta_{ads}H - T\,\Delta_{ads}S < 0$$

where G = Gibbs free energy, H = enthalpy, T = absolute temperature, S = entropy and Δ_{ads} = net change of the thermodynamical parameters. This approach is therefore based on critical surface tension or surface free energy (sfe, mJ/m^2), which is a physical value specifying the whole energy of a solid surface as an equivalent to the surface tension of a fluid (Hannig et al. 2009). This describes interactions at small distances without the influence of repulsive forces and is considered a predominant force for the initial adhesion of cells to solid surfaces (Krekeler et al. 1989; Scheie 1994).

The classical DLVO model (named after Derjaguin, Landau, Verwey and Overbeck) describes the Gibbs energy that is needed for adhesion, which can be calculated as a function of separation distance (Strevett et al. 2003; Vadillo-Rodríguez et al. 2005).

However, since acid-base interactions were neglected in the classical DLVO theory, an extended DLVO theory was described by van Oss to incorporate the thermodynamic aspects. Only the extended DLVO theory could explain empirical data, such as the finding by Absolom et al. (1983) that bacteria with high hydrophobicity adhered more strongly than those of low hydrophobicity.

Factors relating to the habitat, including pH, ion concentration and nutrient availability, also influence the attachment of cells since the ζ-potential, the electrostatic charge of bacterial cells, can be dependent on the ionic strength of the surrounding medium (van Loosdrecht et al. 1987; Pereni et al. 2006; Palmer et al. 2007). However, since Gram-negative species were found to be influenced by the moisture levels of their imminent surroundings, while Gram-positive organisms were less susceptible to influences from outside the cell, due to being shielded by their many peptidoglycan layers, this influence does not seem to be uniform (Strevett et al. 2003). Although research has been focusing on this part of the biofilm formation for some time, some of the mechanisms, especially the transition from reversible to irreversible adhesion, are still not well understood and subject of ongoing research (van der Mei et al. 2008).

1.5.2 Irreversible adhesion

Following these non-specific interactions, bacterial adhesion can become irreversible through covalent, ionic or hydrogen bonding. During the process of bacterial adhesion to oral surfaces, pellicle components such as proline-rich proteins and glycoproteins are vital as they act as specific receptors (Kolenbrander et al. 1993; Marcotte et al. 1998; Mager et al. 2003; Aas et al. 2005). Alongside these pellicle components, bacteria have developed cell appendages to serve as specific adhesins facilitating colonisation of the specific niche in which their target receptors can be found (Scannapieco 1994; Whittaker et al. 1996; Davey et al. 2000; Mager et al. 2003; Marsh et al. 2006). *Streptococcus mitis*, *S. infantis* and *Neisseria subflava*, for instance, can be found predominantly on the hard palate, while other streptococci (including *S. sanguinis*, and *S. gordonii*) favourably colonise the tooth surface (Aas et al. 2005). *S. salivarius*, however, has been found on both buccal and tooth surfaces (Gibbons 1989). *S. mutans* and *S. sanguinis* show a particularly strong predilection for the colonisation of enamel and are not found in the oral cavities of infants prior to tooth eruption, also

disappearing from the mouth following the loss of all teeth (Gibbons et al. 1988; Gibbons 1989).

One of the predominant groups of salivary proteins in the pellicle involved in adhesion are the (acidic) proline-rich proteins (PRPs). Gibbons et al. (1988) showed that *Actinomyces viscosus* (now known as *A. oris* (Henssge et al. 2009)) type 1 fimbriae interact specifically with several proline-rich proteins, binding strongly to them. However, both *A. viscosus* and *S. gordonii* can bind PRPs only in a configuration that is revealed when these proteins are selectively adsorbed to hydroxyapatite (Gibbons et al. 1988; Gibbons et al. 1991). For example, statherin, a salivary pellicle protein with a high content in tyrosine, proline and glutamic acid, can also only be bound by type 1 fimbriae of *A. viscosus* when it is attached to hydroxyapatite (Schlesinger et al. 1977; Gibbons et al. 1988).

Another important group of salivary pellicle proteins involved in the adhesion of microorganisms are salivary glycoproteins, which can be divided in two groups; the first comprising mucous glycoproteins that contain mostly O-linked oligosaccharides and the second group, the serous glycoproteins, containing N-linked oligosaccharides (Levine et al. 1987). Members of this family of salivary glycoproteins comprising the two groups include mucins, proline-rich glycoproteins, α-amylase, lactoferrin and sIgA to name but a few (Levine et al. 1987).

Mucins appear in two distinct chemical forms, a high-molecular-weight mucin glycoprotein 1 (MG1) and a low-molecular-weight mucin glycoprotein 2 (MG2) made up of two isoforms, MG2a and MG2b (Prakobphol et al. 1982; Loomis et al. 1987; Reddy et al. 1992). Both types of mucins have been implicated in bacterial adhesion to the pellicle as described, for example, by Veerman et al. (1995) who found that only *Haemophilus influenza*e out of a pool of oral microorganisms binds MG1 specifically. MG2 on the other hand can promote the adherence of *S. sanguinis,* as well as its aggregation (Ligtenberg et al. 1992). A study examining the interaction of MG2 with *Staphylococcus aureus* and *Pseudomonas aeruginosa* found that MG2 in human submandibular-sublingual saliva bound to the bacterial surface, but the purified isoforms MG2a and MG2b did not. It was subsequently established that only the formation of a MG2-secretory Immunoglobulin A (sIgA) complex was capable of

agglutination of the two pathogens (Biesbrock et al. 1991). The specificity of these interactions was further illustrated by Murray et *al.* showing that two strains of *Streptococcus sanguinis* interact with MG2 in a strain-specific manner, since the removal of sialic acid from MG2 abolished the adherence of only one of the two strains (Murray et al. 1992).

Binding to these host pellicle proteins by bacteria occurs via a group of cell surface proteins called adhesins. Many of the bacterial adhesins, like FimA (Oligino et al. 1993), ScaA (Andersen et al. 1993), ScbA (Correia et al. 1996), PsaA (Sampson et al. 1994) and SsaB (Ganeshkumar et al. 1988) belong to the lipoprotein receptor antigen I (LraI) family of lipoproteins and are often part of ATP-binding cassette (ABC) transporter systems that span the cell membrane (Davey et al. 2000). Others, such as the proteins SspA and SspB, members of the antigen I/II family, also have multiple functions and bind to human salivary agglutinin, collagen, and certain *A. naeslundii* strains (Egland et al. 2001). It has even been proposed that SspA uses two different domains for coaggregation; one produces lactose-noninhibitable bonds, the other lactose-inhibitable coaggregations (Egland et al. 2001).

1.6 Sequence of plaque formation

The timeline of colonising events during the development of plaque has been the subject of extensive research over the years and has offered unique insights into the development of this type of biofilm. A study identifying early microbial colonisers using checkerboard DNA-DNA hybridisation reported that the distribution of bacteria in the developing biofilm immediately after cleaning was very similar to that of saliva, though with time, patterns between saliva and the biofilm became distinct (Li et al. 2004). Indeed, the sequence of plaque formation is relatively specific (Kolenbrander et al. 2002) and certain groups of organisms adhere in a sequential manner due to their nature of displayed adhesins and/or interdependencies with other species.

Streptococcus and *Actinomyces* species, for example, are seen as the predominant species in early developing dental plaque, also known as supragingival biofilm (Nyvad et al. 1987; Li et al. 2004; Diaz et al. 2006; Hannig et al. 2007). Both Li et al. (2004) and Hannig et al. (2007) found that streptococci represent up to 25 % of bacteria in the first 2 h of plaque formation, which is likely explained by the ability of various

streptococci to bind a range of pellicle components, such as glycosyltransferases, amylases, PRPs and mucins (Kolenbrander et al. 2002). Streptococci binding to the pellicle then present a nascent surface to which other streptococcal species and *Actinomyces* species can bind (Palmer et al. 2003). Another study investigating the first 24 h of biofilm formation found that streptococci reached a maximum within 8 - 12 h and remained at this level until the 24 h time point (Nyvad et al. 1987).

Other species of bacteria also play a role in early plaque formation. Two studies examining early plaque formation employing different analysis methodologies showed that *Actinomyces* represented up to 50 % of bacteria at this stage (Nyvad et al. 1987; Ramberg et al. 2003). Similar observations were made by Li et al. (2004) in that *Actinomyces* and also species of the purple and green complex (health associated early colonisers) could be detected and their numbers remained stable over the first 6 h of formation. However, as more streptococci attached, a reduction in the proportion of *Actinomyces* was seen over time. The abundance of *Actinomyces spp.* was suggested to be due to the synthesis of neuraminidases by some species, which results in the exposure of galactosyl residues in MG1 proteins of the pellicle, to which *Actinomyces spp*. can bind with their galactosyl-binding adhesins.

As well as the above examples, various studies have reported the detection of numerous health- and disease-associated species of the genera *Gemella, Granulicatella, Neisseria, Prevotella, Rothia* and *Veillonella* at all time points of plaque formation, although at varying levels of detection (Ramberg et al. 2003; Li et al. 2004; Diaz et al. 2006).

Examining later stages in early biofilm formation, a study by Dige et al. (2007) targeting streptococci with FISH probes in biofilms that were allowed to form on glass slabs worn *in situ* over 48 h, showed distinct colonisation patterns after 6 h, namely, single bacteria, bacteria in pairs and clumps of bacteria of varying sizes. Most cells were identified as streptococci forming small chains, but rods and filaments were also observed showing a diversification of cell morphology. A notable increase was observed for a number of microorganisms between 24 and 48 h, but biofilms in all individuals were dominated by streptococci intermingled with single non-streptococcal cells or groups of non-streptococcal cells (Dige et al. 2007). Guggenheim et al. (2001)

examining a mature biofilm formed over 64 h using a five species *in vitro* model of supragingival plaque also reported mostly discrete microcolonies at 16 h, but further found densely populated biofilms containing large numbers of microcolonies at 40 and 64 h. By 40 h, microcolonies were of diverse morphology with columnar as well as mushroom-shaped forms being observed, showing how different species result in a specific biofilm architecture with water channels facilitating metabolic networks, as will be discussed below.

Extending the observation time points of early biofilm development beyond the immediate and very early stages, two studies looking at four and seven day old plaque found somewhat contrasting results, possibly due to the different methodologies used. Ramberg et al. (2003) looked at microbiota over four days in plaque scraped off from teeth of ten volunteers using a checkerboard DNA-DNA hybridisation assay. Al-Ahmad et al. (2007), on the other hand, examined plaque formed over seven days on bovine enamel slabs worn in situ, assessing the in vivo dynamics of *Streptococcus* spp., *Actinomyces naeslundii*, *Fusobacterium nucleatum* and *Veillonella* spp. using multiplex fluorescence in situ hybridisation. While Al-Ahmad et al. (2007) found the proportion of streptococci over the 7 day period to vary and a decrease in numbers became only significant after seven days, Ramberg et al. (2003) noted a significant increase of *S. gordonii*, *S. oralis*, *S. sanguinis* and *S. anginosus* over four days. Similarly, Ramberg et al. (2003) found no significant change in *Actinomyces* species over four days, whilst Al-Ahmad et al. (2007) found *A. naeslundii* to make up 7.7 % on day 1, decreased significantly on day 2, increased again, but decreased until day 7, when the decrease became again significant. Solely findings regarding the proportion of *F. nucleatum* were similar in that Al-Ahmad et al. (2007) first observed a decrease after two days with a subsequent significant increase at seven days, while Ramberg et al. (2003) noted a significant increase in three *F. nucleatum* phylotypes as well as *F. periodonticum* at four days. Detection levels of *F. nucleatum* in these studies are in agreement with the theory that this organism plays a vital role in the progression of biofilm maturation and can bridge primary and later colonisers (Kolenbrander et al. 2002).

Looking into late-stage plaque colonisation, there are yet further changes in microbial populations. A study of shifts in dental plaque development over nine days into the

late phase of plaque colonisation using culture to determine the presence of several major plaque species including *Neisseria, Nocardia, Streptococci, Fusobacteria, Veillonella, Actinomyces* and *Corynebacteria* also found numbers of *Fusobacteria* to increase over time (Ritz 1967). Since the study was done in the 1960s, when modern molecular analysis techniques did not exist and culture media were not as refined, we must take care interpreting these results, since this study reports *Streptococci, Neisseria* and *Nocardia* to be high initially. On the other hand, a study by Nyvad et al. (1987) two decades later reported only six out of more than 1700 isolates with characteristics of *Neisseria*. Indeed, the authors of this study declared that some proportion of the flora was unaccounted for in all samples and admitted that this was probably due to limitations in the culture techniques used. However, the general tenor of findings mirrors those of more recent studies in that early stages of biofilm formation are characterised by coccal morphology, while levels of filamentous organisms such as *Actinomyces, Corynebacterium* and *Fusobacterium* become significant at later stages (Ritz 1967).

Following assessment of microbial complexes in supragingival plaque over a period of seven days, Haffajee et al. (2008) reasoned that their data indicated seven days is probably not a sufficient time span for full development of the final climax community. This hypothesis makes a study of the development of plaque on epoxy resin crowns using electron microscopy spanning two months of biofilm formation from the 1970s particularly important (Listgarten et al. 1975). This study, too, found primarily coccal forms of varying size and in varying arrangements, as well as few branching filaments at 24 h post cleaning. Little change between 24 h and 72 h was noted with cocci remaining predominant, whilst there were few rods and filaments. It was observed that volunteers who showed little plaque formation at 24h continued with this trend throughout and at seven days these subjects still showed a predominance of cocci (Listgarten et al. 1975). Nevertheless, most subjects exhibited a mixed plaque flora with cocci as well as rods and filaments by seven days. It appeared that filaments were colonising as a top layer over the coccal plaque and went on to invade and subsequently replace the coccal organisms with a primarily filamentous flora. By 21 days of plaque maturation, a shift from mixed cocco-filamentous plaque near the gingival sulcus to a predominantly filamentous plaque was observed. However, the

closer to the crown, the more plaque showed features of earlier time points of plaque maturation. As would be expected, the two month samples provided the most voluminous plaque. The bulk of the biofilm at this time was made up of densely packed filamentous microorganisms oriented perpendicular to the crown surface, while the most superficial layer of the plaque near the junctional epithelium was composed of a fuzzy layer of thin bacterial cells including numerous spirochetes.

Since that study, not much has been done to further define the final phases of plaque biofilm formation. The only recent study looking at biofilm architecture of late-stage / mature biofilms (of undetermined age) used fluorescence in situ hybridisation to localise the most abundant species associated with periodontal disease in supra- and subgingival biofilms (Zijnge et al. 2010). Zijnge et al. (2010) identified that supragingival biofilms were generally more heterogenous in architecture than subgingival biofilms. Additionally, the structures of the biofilms were divided into layers, of which the basal layer was further classified into four types of biofilms:

i) rod-shaped *Actinomyces* that were oriented perpendicular to the tooth surface,
ii) a mixture of *Actinomyces* and chains of cocci (not identified as streptococci) that were perpendicular to the tooth surface,
iii) filamentous bacteria, streptococci and yeast, where streptococci formed a distinct colony around the yeast cells, and
iv) predominantly streptococci growing in close proximity to *Lactobacillus* spp., again perpendicularly to the tooth surface.

A second layer, which could be found on any of the above mentioned four types of biofilm, covered the basal layer, and streptococci were present either as heterogeneously scattered cells without any apparent organisation, or as a thin coat. It was speculated that the presence of *Streptococcus* sp. or members of the CFB-cluster (*Cytophaga, Flavobacterium* and *Bacteroides*) in the second layer could reflect the crucial transition from predominantly Gram-positive saccharolytic plaque bacteria to Gram-negative proteolytic plaque bacteria, possibly due to availability of nutrients (Zijnge et al. 2010).

This idea of a shift in the plaque population over the course of time and between individuals is not new. Indeed, it was suggested that the transition from an aerobe to an anaerobe microclimate may be facilitated by aerobic *Neisseria*, aerotolerant streptococci and *Gemella*, which formed part of a core group in all subjects in a study characterising the subject-specific oral microbiota during initial colonisation of enamel (Diaz et al. 2006). Indeed, inter-dependencies between species are established very early in the colonisation process, which can be exemplified by *Veillonella* that bind to streptococci and establish a food network by utilising lactic acid, a streptococcal metabolic end product (McBride et al. 1981).

1.7 Bacterial interactions in the process of biofilm formation

The process of adhesion between two genetically distinct bacterial species, as described for streptococci and *Veillonella* species, is called coaggregation (Kolenbrander et al. 1993; Kolenbrander et al. 2006). Coaggregation is a non-random, highly specific inter-species interaction mediated by lectin-carbohydrate cell surface structures. These surface components consist mostly of proteins on the participating cells that are protease-sensitive and fimbriae-associated and which can be heat-inactivated (Kolenbrander et al. 1986; Weiss et al. 1988; Kolenbrander et al. 1989; Kolenbrander et al. 2005). An example of a typical carbohydrate-lectin interaction is the lactose-inhibitable *Fusobacterium nucleatum-Porphyromonas gingivalis* coaggregation (Kinder et al. 1993). This bond is made up of a protein adhesin on the *F. nucleatum* cells and a cognate carbohydrate receptor on the *P. gingivalis* cells (Kinder et al. 1993). The carbohydrate receptors, complementary to the *F. nucleatum* expressed adhesin, are heat-stable and protease-insensitive (Kolenbrander et al. 2005). Carbohydrate receptors are ideal as a recognition system, since they can combine in so many ways, thus containing huge amounts of molecular information. They are, however, not the only aggregation system utilised by microbes. Proteinaceous receptors to which adhesins bind stereochemically have also been discovered (Gibbons 1989).

Guggenheim et al. (2001) reported that intraspecies aggregation and coaggregation were not involved in initial plaque formation, since 15 minutes after professional cleaning, the majority (90 %) of cells adhered to the salivary pellicle were non-aggregated cells (Guggenheim et al. 2001). However, findings of many studies contest

this and report coaggregation as an important factor in early colonisation of oral surfaces (Kolenbrander et al. 1990; Palmer et al. 2001; Palmer et al. 2003; Diaz et al. 2006; Hannig et al. 2007).

The reasons for microbial coaggregation are many but predominantly represent a mechanism to allow microbes to occupy otherwise hostile environments. Because the nutrients available to bacteria in the oral cavity are complex and each oral species possesses, at most, only a few of the enzymes needed to metabolise the substrate, the lifestyle of a mixed species community is advantageous for all members participating in the nutrient degradation (Kolenbrander et al. 2005). For some species, coaggregation is essential for survival as can be shown for *Streptococcus gordonii* and *Actinomyces naeslundii*. These two microbes flourish in the oral cavity when they interact and coaggregate, but when grown individually in a saliva-coated flow cell neither species survives (Egland et al. 2001; Palmer et al. 2001).

A second feature of coaggregation presents itself when two different species can attach to a third species by different means (Kolenbrander et al. 1993; Guggenheim et al. 2001). This principle of sequential coaggregation is called the principle of bridging, and strains from multiple species can thus recognise an identical receptor on a common partner bearing two or more receptor polysaccharides, or various types of adhesins, or a mixture of the two, allowing three or more microorganisms to be brought into close proximity (Kolenbrander et al. 2006). *F. nucleatum* represents an important species in this process, bridging the sequence of early and late colonisers by binding both, despite late colonisers generally not being able to bind primary colonisers (Kolenbrander et al. 2002).

Coaggregated microcolonies also display different morphological shapes such as rosettes (Kolenbrander et al. 1988) and bristly brushes (Robert et al. 1990), which are determined by the participating cell types and ratios. *Fusobacterium* species can, for example, coaggregate with streptococci and *Veillonella atypica* PK1910, and corn-cob structures will form in both instances (Kaufman et al. 1989; Kolenbrander et al. 1993). Rosettes will form if small cells like streptococci are mixed with rod shaped cells, like bacteroides, in a 10:1 ratio (Kolenbrander et al. 1988). The exterior cells of these rosettes can also function as bridging organisms, binding exterior cells of other

rosettes and thereby building large multigeneric coaggregates (Kolenbrander et al. 1993).

1.8 Biofilms

The description of plaque so far has been based on the actual microbes and how adhesion to surfaces and / or other cells is achieved, as well as the sequential order of bacterial attachment in the developing supragingival plaque. However, dental plaque is just one example of a specific biofilm and generally biofilms, be it dental plaque or biofilms on rocks in rivers or on ship hulls, are made up of much more than just the microbes. Furthermore, its components fulfil important structural and protective functions.

Biofilms are matrix-enclosed bacterial communities in which the microorganisms live in symbiotic and/or syntrophic relationships in a surface-attached community (Costerton et al. 1995; Allison 2003) as opposed to a planktonic lifestyle. Symbiosis, the association of organisms in a community in which at least one party of the relationship profits, and syntrophy, a metabolic symbiosis, characterise most of the relationships between cells in biofilms (Lopez-Garcia et al. 1999; Moran 2006).

1.8.1 Biofilm structure

The biofilm matrix, also called the glycocalyx, that envelopes the adherent bacteria of a biofilm consists of a rich blend of different molecules. It contains mainly water (up to 97 %) but also consists of a mixture of exopolysaccharides (EPS), nucleic acids, proteins, glycoproteins and phospholipids (Sutherland 2001; Allison 2003). Studies measuring the total organic carbon of biofilm matrices have suggested that cellular material actually represents as little as 2-15 % of these structural/solid elements, while the proportion of exopolymeric substances can make up to 88 % of the structural matrix components (Lawrence et al. 1998; Sutherland 2001).

Of these components exopolysaccharides are the most important structure-giving ones, and their composition is influenced by many intrinsic and extrinsic factors such as the genotype of the bacterial cells and physico-chemical properties of the surrounding environment (Christensen et al. 1985; Allison 2003). As shown in the case of *P. aeruginosa*, sessile cells in the process of forming a biofilm show a different protein expression pattern as compared to planktonic cells (Svensäter et al. 2001;

Welin et al. 2004). Many genes are specifically up or down regulated when the bacteria are in a sessile state, and some of these expressed proteins have structure-giving functions (Svensäter et al. 2001; Welin et al. 2004). For instance, the staphylococcal polysaccharide intercellular adhesin (PIA), expressed by *Staphylococcus epidermidis* and *Staphylococcus aureus*, and its related polymer poly-N-acetylglucosamine (PNAG), both serve as adhesins and are required for biofilm formation (Branda et al. 2005).

The exopolysaccharides making up the glycocalyx can be subdivided into two groups - homopolysaccharides and heteropolysaccharides, each of which have different structures and therefore influence the physico-chemical properties of the matrix. Homopolysaccharides exist either as linear molecules with a single linkage type, as linear units possessing a one-sugar side chain or as branched structures (Duenas-Chasco et al. 1998), whereas heteropolysaccharides consist of repeating backbone units varying in size from disaccharides to octosaccharides (Gruter et al. 1993; Robijn et al. 1996). The matrix composition has a fundamental impact on the architecture, since proteins in the matrix can directly or indirectly cross-link, thereby enhancing the structural complexity and giving mechanical stability (Sutherland 2001; Allison 2003).

Another important structural feature of biofilms are water channels, which present a very important aspect in the biofilm architecture, since they not only influence the configuration, but also provide the bacteria deeper within the structure with oxygen and/or nutrients. The water channel system in biofilms has even been compared to a primitive circulatory system analogous to higher organisms in which not only are nutrients delivered to the many micro-niches, but metabolic products are also removed at the same time (Costerton et al. 1994). Although water channels can distribute oxygen into the biofilm, diffusion limitations and oxygen utilisation by bacteria mean that towards the centre and base of each biofilm the number of channels decreases and oxygen levels are low (Costerton et al. 1995; Robinson et al. 2006). This provides a likely explanation for the fact that fastidious anaerobes can grow in close proximity to aerobes and suggests that the spatial composition of biofilm communities develops in a unique way depending on the substrate, available nutrients and appearance of microorganisms (Costerton et al. 1995; Kolenbrander 2000). For example, *P. aeruginosa*'s rhamnolipid production (a process that is under quorum sensing control) has an important function for the shape of the biofilm, since it keeps

the water channels open during the maturation of the matrix (Davies et al. 1998; Branda et al. 2005). Consequently, one of the most important functions of the biofilm matrix is the prevention of desiccation, protecting not just the bacterial cells, but also metabolic networks within (Chang et al. 2007).

Three conceptual models describing the architecture of biofilms have been suggested. The first model, the heterogeneous mosaic biofilm model proposed individual stacks of microbial cells that are unconnected and surrounded by water (Wimpenny et al. 1997). The second model known as the penetrated water-channel biofilm model, is characterised by microcolonies forming mushroom-like structures that can merge at the top (Wimpenny et al. 1997). Finally, the third model, the dense confluent biofilm, describes cells growing several layers high, without any water channels or other structure-giving features. It is currently thought that most biofilms are a conglomerate of two or all three models and consist of a more dense biomass at the base with water channels of varying sizes, depending on the extrinsic conditions (Robinson et al. 2006). It is important to note, however, that the reported observations of dense biofilms could be due to preparation artefacts of the transmission electron microscope that was used at the time (Wood et al. 2000).

During the life of a biofilm the mean proportion of viable cells averages around 80 %, as could be shown by live/dead stains (Guggenheim et al. 2001). The 20 % of non-viable cells provide a source of extracellular DNA as well as cell detritus during biofilm formation that can be incorporated into the biofilm matrix (Branda et al. 2005). It has been shown, for example, that *Pseudomonas aeruginosa*, which is known for its extracellular DNA excretion via vesicles, was not able to produce a biofilm in the presence of DNase I (Whitchurch et al. 2002; Nemoto et al. 2003). Even mature biofilms of up to 60 h old were degraded by the enzyme, which gives a strong indication that extracellular DNA is essential for the establishment of *P. aeruginosa* biofilms and may play a role in the biofilm formation process of other bacteria that also release DNA (Whitchurch et al. 2002). Since then, more studies have examined this hypothesis using strains of *Streptococcus* (Petersen et al. 2005), *Staphylococcus* (Qin et al. 2007) and *Bordetella* (Conover et al. 2011). All of these studies confirmed the importance of DNA as a matrix constituent. A recent study by Das et al. (2010)

further showed that extracellular DNA from Gram-positive bacteria plays an important role in initial adhesion and aggregation of bacteria on surfaces.

1.8.2 Matrix functions

An important function of the glycocalyx is protection of the biofilm against destructive forces of many different kinds (Welin et al. 2004). The matrix offers some resistance to protozoal grazing and host defence mechanisms such as phagocytosis (Weitere et al. 2005). *Porphyromonas gingivalis*, for example, can release lipopolysaccharides (LPS) or LPS-containing outer membrane vesicles that saturate antibodies on or just below the surface of the biofilm matrix before they can reach the bacterial cells within. Moreover, a recent study explored how *S. mutans*-produced EPS-matrix modulates biofilm architecture and population shifts. It was observed that *S. mutans* gtfB/gtfC genes, whose expression was enhanced in the presence of *A. naeslundii* and *S. oralis*, mediated the creation of compartmentalised acidic and EPS-rich microenvironments. *S. mutans* can proliferate in these microenvironments, thereby posing an essential factor associated with virulence in cariogenic biofilms (Xiao et al. 2012).

Another equally important defence mechanism is protection against and resistance to antimicrobial agents. One of the first hurdles that antibiotics have to overcome, for example, is the matrix itself and its associated diffusion limitation and this also pertains to host anti-microbial peptides. The diffusion reaction theory states that the biofilm structure may restrict the penetration of compounds, and charged molecules can be intercepted by matrix polymers (Marsh 2003). The EPS was found to show characteristics similar to an ion exchange resin and can actively remove strongly charged molecules such as antimicrobial agents from the surrounding solution, leaving the inner cells of the biofilm unharmed (Wolfaardt et al. 1998; Gilbert et al. 2001).

It has long been known that higher concentrations of antibiotics are needed to eradicate a biofilm, and Johnson et al. (2002) reported that for all strains tested in their experiments, biofilms were two- to eight-fold less susceptible to the biocide para-chloro-meta-xylenol (PCMX) than planktonic cells of the same strains. This could be due to a slower growth rate of cells living in a biofilm. It is also known that attached cells, and therefore biofilm cells, show a distinctly different protein expression pattern. This altered regulation of gene expression could lead to the development of a bacterial

community with specific growth and metabolic properties in which antibiotics cannot find a contact point (Svensäter et al. 2001). Boles et al. (2004) suggested the so-called insurance hypothesis, which relates to the effects of diversity on a community as a whole. The researchers found in their experiments that the mode of biofilm growth induces variants in subpopulations that have specialised biofilm functions, which ensure the survival of the community as a whole with changes in environmental conditions.

1.9 Dental caries

Dental caries is the progressive decalcification and subsequent destruction of the organic matrix of the tooth as a result of metabolic processes of bacteria growing in a plaque biofilm attached to the tooth surface.

1.9.1 Structure of enamel and dentine

The tooth is a complex structure made up of distinct structures of enamel, dentine and the pulp (Figure 1.1). It is anchored in either the mandible or maxilla and periodontium by its cementum-covered roots and periodontal ligament, which connect the tooth to the blood circulatory and nervous system. In health only the enamel-covered crown is exposed to the oral cavity. Two of these structures (enamel and dentine) are comprised of a unique material only found in the tooth.

1.9.1.1 Enamel

Enamel is the most highly mineralised tissue in the human body consisting of 96 % mineral (hydroxyapatite) and 4 % organic material and water. It differs from other mineralised tissues such as bone or dentine in that it is non-collagenous (Fincham et al. 1999; Nanci 2008). It is a non-vital, relatively brittle tissue of epithelial origin that cannot regenerate and the hydroxyapatite crystals' susceptibility to acid dissolution is the basis for the development of dental caries and erosion (Fincham et al. 1999; Nanci 2008).

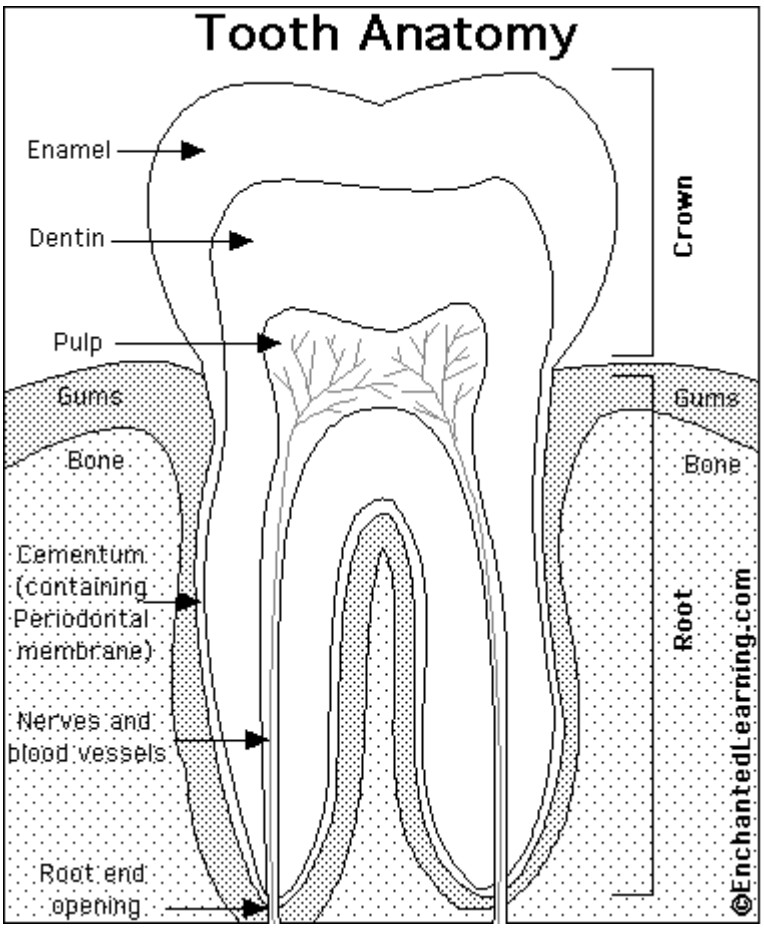


Figure 1.1: Schematic representation of the tooth anatomy (source: http://www.enchantedlearning.com/subjects/anatomy/teeth/toothprintout.shtml)

1.9.1.2 Dentine

The underlying dentine, which makes up the bulk of the tooth, supports enamel and compensates for its brittleness. Dentine is made up of about 70 % inorganic material, 20 % organic material and 10 % water by weight (45 %, 33 % and 22 % by volume, respectively) (Nanci 2008). The inorganic phase consists of hydroxyapatite, and the organic phase consists of approximately 90 % collagen. The collagen consists predominantly of type I and small amounts of types III and IV. The collagen is organised in fibrils that are evenly spaced, linearly aligned groupings of molecules, which store more than 50 % of the mineral in its holes and pores (Banerjee 1999; Nanci 2008). The non-collagenous proteins (NCP) consist of dentine sialoprotein, dentine phosphoprotein, bone sialoprotein, dentine matrix protein-1, osteoprotein and matrix extracellular phosphoglycoprotein (Smith et al. 2012). Phosphorylated phosphoproteins make up the majority of non-collagenous proteins, and these NCPs

have been shown to regulate mineral deposition (Clarkson et al. 1998; Suzuki et al. 2009).

1.9.1.2.1 Types of dentine

There are three different stages of dentine formation – primary, secondary and tertiary. While the actual tooth structure consists of primary dentine, the other two types are laid down as a consequence of the aging process and external noxious stimuli, respectively.

Primary dentine, which can be subclassified according to amount of mineralisation and location, represents the major part of the tooth and is produced during tooth formation. During dentinogenesis, the formation of dentine, odontoblasts produce a 10 - 30 µm wide unmineralised zone of dentine closest to the pulp, which is known as predentine. Predentine consists primarily of collagen I, but is converted to dentine by gradual mineralisation through incorporation of NCPs (Butler 1995; Nanci 2008). Primary dentine can be subclassified into mantle dentine, which represents the first layer of primary dentine produced directly subjacent to the enamel at the enamel-dentine junction (EDJ); and the remaining dentine known as circumpulpal dentine. Circumpulpal dentine can, in turn, be subdivided into intertubular and peritubular dentine with respect to its relative relationship to the dentine tubules. The matrix of dentine consists of tightly packed dentine tubules, which extend through its entire length/thickness. These tubules are surrounded by the highly mineralised peritubular dentine, which in turn is embedded in the intertubular dentine. The number of tubules varies from about 45000 / mm^2 near the pulp to about 20000 / mm^2 near the EDJ (Garberoglio et al. 1976). The diameters of tubules of human teeth vary between 1 - 2.5 µm (Garberoglio et al. 1976). Dentine tubules have been found to create a branching pattern resulting in an anastomosing system through the intertubular dentine (Mjor et al. 1996), thus allowing bacterial products, or bacteria themselves, to diffuse and/or invade the tubules in the event of exposure through trauma and/or caries (Love et al. 2002).

Secondary dentine represents the sustained, but much slower, deposition of dentine once root formation has been completed (Burke et al. 1995). Secondary dentine has a

tubular structure like primary dentine, and is laid down as a continuation of the former (Nanci 2008).

Tertiary dentine is produced in reaction to a stimulus or injury and was described by Stanley et al. (1966) as being localised exclusively adjacent to the irritated zone, its tubules being very irregular, tortuous, and reduced in number or even absent. Tertiary dentine can be subclassified as reactionary dentine, which is deposited by pre-existing odontoblasts or reparative dentine, which is formed by newly differentiated odontoblast-like cells (Nanci 2008). It was found the more severe the trauma, the more rapidly dentine was produced (Stanley et al. 1966).

1.9.2 Clinical aspects of caries

Damage to enamel and/or dentine by microbes resulting in caries is the most common form of tooth damage and has specific, well-defined pathologies. Dental caries lesions can be characterised using various criteria. Anatomical sites can be occlusal, meaning the lesion develops in the pits and fissures of the crown of molars, which is also known as coronal caries. Caries attack on smooth surfaces can lead to proximal (between teeth) and cervical lesions. Root caries can develop when the cementum becomes exposed due to periodontal disease, old age or too vigorous tooth brushing. Lesions characterised as simple involve one tooth surface, while compound lesions involve two surfaces and complex lesions involve three or more. An evidence-based clinical caries assessment system known as the modified International Caries Detection and Assessment System (mICDAS) allows characterisation of tissue involvement according to five numbered subdivisions: 0 stands for no or slight change in enamel translucency after prolonged air drying (> 5 sec) with no enamel demineralisation. A white spot lesion that becomes visible after air drying and enamel demineralisation limited to the outer 50 % of enamel are represented by 1, while a white spot lesion that is visible without the need for air drying combined with demineralisation involving the inner 50 % of enamel through to the outer third of dentine is represented by 2. A 3 stands for the localised breakdown of enamel that is accompanied by a greyish discolouration from the underlying dentine in some cases. Here the demineralisation reaches as far as the inner third of dentine. Lastly, 4 describes the gross cavitation exposing the underlying stained dentine, in which the demineralisation progresses towards the pulp (Banerjee et al. 2011). Lesion progression is classified as active or arrested. Active,

rampant lesions, where the process of cavity formation is accelerated and which are associated with a highly cariogenic diet, progress rapidly and usually many teeth are affected. In contrast, arrested lesions have stopped progressing and are inactive (Banerjee et al. 2011). Because the advancing lesion/dissolution of enamel matrix is influenced by enamel prisms, the shape of the lesion resembles that of a cone, with its base toward the enamel-dentine junction (Kidd et al. 2004).

1.9.3 Clinical lesion development

Caries develops due to bacteria producing acids when metabolising fermentable carbohydrates, resulting in demineralisation of the tooth matrix. This demineralisation is not uniform as hydroxyapatite in enamel and dentine is not pure. About 1 in 10 phosphate ions (PO_4^{3-}) in enamel and about 1 in 5 in dentine are substituted by carbonate ions (CO_3^{2-}), resulting in defects and calcium deficient regions that make these regions more acid soluble than pure hydroxyapatite would be (Featherstone 2008). When acids reach these susceptible sites by diffusion, calcium (Ca^{2+}) and PO_4^{3-} are dissolved into the surrounding aqueous phase. A study of the development of subsurface enamel lesions during demineralisation using x-ray microtomography, provided evidence that local variations in fractional pore volume of partially demineralised enamel influence the subsequent spatial development of the lesion (Dowker et al. 2003). Lesion morphology was found to be consistent with preferential anisotropic dissolution following the prism direction at the advancing front.

Despite the development of these lesions, dissolution of tooth matrix is not irreversible at this stage remineralisation can occur, especially if fluoride ions (F^-) are present in sufficient concentrations (Lynch et al. 2004). F^- adsorbs to the crystal surface by attracting Ca^{2+} ions, which are held in a supersaturated state by salivary proteins such as statherins, making them readily available. PO_4^{3-} ions are in turn attracted by the Ca^{2+} ions and the three components start to build a fluorapatite-like remineralised veneer on the crystal surface, which is less soluble than the original mineral (Amjad et al. 1979; Featherstone 2008). This process provides an explanation for the fact that hyposalivation can result in increased incidence of caries as in this case, insufficient Ca^{2+} and PO_4^{3-} are available for remineralisation.

The next stage in the progression of enamel caries is the development of a histological translucent zone in which porosity is increased by approximately ten times, which enlarges in time with a dark zone in its centre (Soames et al. 1993). The continued demineralisation of enamel results in the centre of the dark zone to become the body of the lesion, at which point the lesion is clinically recognisable as a white spot lesion – the first, macroscopically visible stage of dental caries. They become visible due to increased enamel porosity following the demineralisation process. The irregular surface generated by the erosion of the outermost surface results in an altered reflection of light compared to the unaffected, comparatively shiny enamel (Fejerskov et al. 2003). Despite this, even if a lesion has progressed to the state of a white spot lesion, the arrest of the lesion progression and remineralisation, at least to an extent, is possible (Dirks 1966).

If it progresses, the lesion starts to spread along the EDJ and a cavity forms due to breakdown of the surface zone (Soames et al. 1993). When the carious process progresses into the dentine, lesions can be divided into zones described as infected dentine, affected dentine, the hypermineralised translucent zone and secondary dentine towards the advancing front approaching the pulp chamber (Banerjee et al. 2011). The translucent zone was found to extend at the advancing front of the carious lesion, building up a defensive wall of hypermineralised dentine towards the pulp chamber (Arnold et al. 2003).

1.9.4 Microbiology of caries

Although the aforementioned 'legend of the toothworm' was described as early as 7000 years ago, it took until the 17th century and the first crude microscope to realise that dental plaque contained bacteria. It was much later before dentistry was established as a field in its own right and researchers started investigating the pathogenesis of caries. W. D. Miller was arguably the first oral microbiology researcher and made important contributions to the field in the late 19th century. He believed that 'both acids and fungi are concerned in producing caries' and formulated the chemico-parasitic theory (Suddick et al. 1990).

However, Miller does not seem to have linked plaque with the occurrence of dental caries, a connection that was only made by G. V. Black and J. L. Williams, who

described the "gelatinous microbic plaques", and who both agreed that caries was due to attack from acids produced by bacteria in these plaques (Suddick et al. 1990).

Bacillus acidophilus (now *Lactobacillus acidophilus*) was identified as a first caries-associated microorganism by Howe in 1917. However, Clarke (1924) did not find *B. acidophilus* regularly present in early caries lesions, and instead isolated *Streptococcus mutans* regularly from lesions in the early stages. Despite this, *L. acidophilus* remained a focal point of research and was believed to be the main causative agent of tooth decay (Jay 1938). In the 1960s, experiments with germ-free rats inoculated with bacteria from a conventional rat and fed a high sugar diet confirmed that caries is indeed a bacteria-associated disease (Fitzgerald et al. 1966); and experiments with streptococcal strains from lesions in rats and hamsters provided proof that these species can cause caries when transmitted into caries-resistant rats and hamsters (Hamada et al. 1980). Subsequent studies managed to isolate streptococci from humans and infect germ-free animals, lending additional support to the importance of these species in the infection process (Krasse 1966; Hamada et al. 1980).

In the late 1960s, more than four decades after Clarke initially described the detection of *S. mutans*, Carlsson reported that those streptococcal strains isolated from caries lesions in humans had similar properties to those described by Clarke and research efforts were concentrated on this organism for decades to follow (Loesche et al. 1975; Hamada et al. 1980; Kohler et al. 1981; Loesche 1986). What is more, studies were even directed at the possibility of a vaccine against *S. mutans* for caries prevention (Lehner et al. 1975; Lehner et al. 1985).

It should be noted, however, that lesions free of *S. mutans* have been observed (Marsh et al. 1989; Kanasi et al. 2010; Lima et al. 2011; Tanner et al. 2011) and other bacteria such as the so-called “low pH” streptococci, *Actinomyces* and *Bifidobacterium* have been shown to be highly acidogenic (van Houte et al. 1996; van Ruyven et al. 2000) and may contribute to the demineralisation of the tooth.

For the assignment of a causative agent of dental caries the aforementioned specific-, non-specific and ecological plaque hypotheses were developed, and the ecological plaque hypothesis is now widely accepted as the most likely scenario, being supported by recent studies. Nevertheless, an extension and complementary in-depth

explanations have been put forward over the last decade. Kleinberg (2002), for example, suggested an alternative aetiological hypothesis. It was argued that lesions can develop in the absence of lactobacilli and other acidogens can provide the necessary acid, which means these species can, at most, only be shown to have an association. They do not, however, represent the sole aetiologic factor. Likewise, mutans streptococci were identified as a possible caries-causing candidate due to associations with availability of fermentable carbohydrates in the diet and prevalence of dental caries, but again, this proved association, not a sole aetiologic relationship. Kleinberg (2002) referred back to the Stephan curve, a characteristic measure of the change in dental plaque pH in response to a challenge. It shows the ability of plaque to produce a rapid and substantial decrease in pH *in vivo* after exposure to fermentable carbohydrate with a subsequent (relatively slow) rise back to resting pH. The study in question showed that pure cultures of many bacteria (at sufficient concentrations) produce pH reductions, but the characteristic curve with the typical ensuing rise of pH observed in plaque *in vivo* could not be reproduced with either pure cultures or mixtures of bacteria (Stephan et al. 1947). In later experiments, Kleinberg (1961) showed a cause-and-effect relationship of the magnitude as well as duration of pH fall due to the bacteria present and substrate availability. Further experiments identified the factor inducing the pH-rise in saliva as mostly arginine made available from small arginine peptides; and finally the addition of arginine and glucose as substrate enabled researchers to reproduce the Stephan curve *in vitro* (Kleinberg 2002).

The conclusion Kleinberg (2002) drew from the described findings above was that since streptococci and lactobacilli are non-arginolytic (non pH-raising), while health-associated bacteria (*S. sanguinis, L. brevis*) are arginolytic (base producing and thereby pH-raising) the deficiency in base formation, which is associated with saliva, can be as important as the formation of acid from fermentable carbohydrates in the process of caries initiation and development. This line of thinking supports the ecological hypothesis and is supported by the fact that hyposalivation is associated with an increased risk of caries formation (Su et al. 2011) and thus a highly attractive hypothesis.

Similar to Kleinberg, Takahashi et al. (2008) reviewed the microbial dynamics of the caries process and proposed an extension of the ecological plaque hypothesis to

explain changes of the phenotypic and genotypic properties of plaque bacteria and the demineralisation and remineralisation balance of the caries process. They stated that from an ecological point of view it is not only important to describe which bacteria are present/involved in caries, but also to know what these bacteria are doing (Takahashi et al. 2008). The suggested extended ecological hypothesis is made up of three building blocks:

The dynamic stability stage: non-mutans bacteria such as non-mutans streptococci (non-MS) and *Actinomyces* dominate in dental plaque and are responsible for the maintenance of dynamic stability. This means that acids are produced from fermentable carbohydrates and subsequent demineralisation of the enamel can occur; however, temporary decreases in pH return to neutral levels by homeostatic mechanisms (Marsh 2006; Takahashi et al. 2008).

The acidogenic stage: if the pH decreases, the acidogenicity and acidurance of non-MS may be enhanced (Takahashi et al. 1999); resulting in an acid-induced selection for low pH non-MS and *Actinomyces* and a subsequent microbial population shift. This, in turn, adversely effects the de- and remineralisation balance, ultimately resulting in lesion initiation (Takahashi et al. 2008).

Aciduric stage: if the balance does not return to its microbial homeostasis and the acidic environment prevails, mutans streptococci (MS) and lactobacilli may replace the low pH-non MS due to being more competitive than the latter under prolonged aciduric conditions (Takahashi et al. 2008; Horiuchi et al. 2009).

Actinomyces and non-MS feature heavily in the explanation of the extended ecological hypothesis, although Takahashi et al. (2008) acknowledge that this is due to detailed studies having been conducted on these organisms and that it is a possibility that other non-mutans aciduric bacteria may be found to be associated with dental caries. They make it clear that it is not the genotype, but the phenotype displayed in a certain environment (i.e. the acidogenic and aciduric potential of involved bacteria) that is fundamental in the potential microbial shift leading to caries.

Data is now appearing that supports this hypothesis. A recent study on microbiota on developing lesions on human enamel over seven weeks using an *in vivo* model and the

Human Oral Microbe Identification Microarray (HOMIM) found that the microbiota of plaque over white spot lesions differed significantly compared to sound enamel (Torlakovic et al. 2012). Twenty-five species, such as the previously caries-associated *S. mutans* and *Lactobacillus* spp., as well as *Atopobium parvulum*, *Dialister invisus*, non-mutans streptococci and others, were found to be associated with initial enamel lesions. These findings are in agreement with studies that found few MS in plaque over white spot lesions, but higher proportions of low pH non-MS or other types of low pH bacteria (Sansone et al. 1993; van Ruyven et al. 2000; Aas et al. 2008). Indeed, Aas et al. (2008) reported that several disease-associated species with distinct bacterial profiles at each stage of caries progression could be found and, more precisely, white spot lesions contained species like *S. parasanguinis* and *S. salivarius* at high levels in both primary and secondary dentitions.

However, other studies assessing microbiota of the different layers of the carious lesion, once the caries lesion had progressed beyond the white spot stage and into dentine, could not confirm these results. Munson et al. (2004) found no difference in the composition of microbiota between the middle and advancing front of the lesion. They did, however, concede that crude sampling might have caused mixing of species. Lima et al. (2011), on the other hand, used a checkerboard DNA-DNA hybridisation analysis, similar to the study carried out by like Aas et al. (2008), but did not observe significant differences for the prevalence of the target bacteria in the three dentine caries layers, with the exception of *Olsenella uli*, which was only found in superficial layers. This difference between the two studies may be due to the fact that Lima et al. (2011) used probes to target 28 bacterial taxa, while Aas et al. (2008) evaluated samples using probes for 110 prevalent bacterial species.

A further study comparing inherent acid tolerance of microbiota of plaque from tooth sites in subjects with and without initial caries found that approximately 1 % of the cultivable taxa grew at the critical pH of 5.5 and below at which tooth enamel demineralises (Svensater et al. 2003). Analysis indicated that MS not only accounted for less than half the streptococcal viable count, but this group was highly variable with respect to acid tolerance, potentially explaining why mutans streptococci are linked to lesion initiation, where acid concentrations are still moderate in comparison to later, advanced stages, but not as much in lesion progression.

Various studies examining advanced lesions are in agreement that certain genera are more predominant in advanced lesions than others, but that a varied microbiota can be found and inter-patient variation can be immense (Munson et al. 2004; Nadkarni et al. 2004; Chhour et al. 2005; Aas et al. 2008). One study for example, found that number of taxa varied from 7 to 31 per patient and that lactobacilli made up 50 % of taxa, while *Prevotella* comprised 15 % (Chhour et al. 2005). In fact, patients could be grouped into *Lactobacillaceae*-dominated or *Prevotellaceae*-dominated lesions, while some patients had a mix of both or very few to none of either. Another study observing a grouping of patients, found groups were dominated by *S. mutans*, *Lactobacillus*, a mix of both or none of either (Gross et al. 2010). These results confirmed findings from a previous study in which genus-specific PCR primers were designed to target lactobacilli in advanced lesions, and lactobacilli were found in all 65 samples (Byun et al. 2004). Species-specific primers for the prominent *Lactobacillus* species found that at least three out of a pool of nine most prominent species were detected in most of the dentine samples (Byun et al. 2004). Discrepancies in frequency of detection in the various studies might be due to usage of universal primers by some studies, while others used specific primers.

After Martin et al. (2002) reported a high frequency and abundance of *Prevotella*-like bacteria in carious dentine, Nadkarni et al. (2004) used *Prevotella*-specific PCR primers to evaluate the occurrence of *Prevotella*-like bacteria in caries and found all detected *Prevotella* species grouped into six clusters, which were represented in most patient samples. The detection of one dominant cluster stood out since these sequences were most closely related to an unidentified rumen bacterium, which was considered to be unrelated to any known cultured *Prevotella* species. Results by Yang et al. (2012) confirmed the importance of *Prevotella* species in relation to caries in their study analysing saliva microbiota from healthy and caries-active subjects. They found the genus *Prevotella* to be significantly associated with caries status, not just with regards to numbers – *Prevotella* were enriched by 10 %, but also with regards to arrays of *Prevotella* species that made up the populations in health and disease. *Prevotella* species could be detected in both healthy and infected individuals, but the distribution of taxa differed in health compared to disease (Yang et al. 2012).

Despite research efforts trying to characterize microbiota responsible for causing dental caries over more than a century, new species and genera are being found continually and novel associations of specific species, such as *Scardovia wiggsiae* (Tanner et al. 2011) are being made. Continual changes are also being made in the methodology used to identify the constituents of the microbiome; for example, Munson et al. (2004) reported that molecular analysis using the 16S rRNA gene underrepresented the phylum *Actinobacteria* and its species with high G+C genomes. Since then, many studies have either focused, or at the very least, facilitated detection of this group by using culture methods and/or molecular analysis methods specifically targeting individual species or genera belonging to the phylum *Actinobacteria*. The aforementioned *S. wiggsiae,* a member of this high G+C phylum was detected by using a culture approach, while a study examining the same sample pool using 16S rRNA PCR/cloning method could not detect any *Bifidobacteriaceae* unless specific PCR primers were used (Kanasi et al. 2010). Employing specific *Bifidobacteriaceae* primers, the detection rate rose to 87 % for children with severe early childhood caries and 21 % for caries-free children. This study even proposed a new caries pathogen candidate from within the family of *Bifidobacteriaceae*. Similarly, Aas et al. (2008) found *Actinomyces* spp. and non-MS were detected at high levels in caries initiation in children and young adults compared to other disease states, while the microbiota of deep dentine caries was dominated by *S. mutans*, *Lactobacillus* spp., as well as the G+C-rich species *Propionibacterium* spp. FMA5 (now *P. acidifaciens*) and *Atopobium* genomospecies C1. Additionally, this study found that *S. mutans* seems to have a more dominant role in dentine and deep dentine caries of primary rather than secondary teeth, while at the same time, occurrence of *S. mutans* was associated with significantly raised levels of species such as *Atopobium* genomospecies C1 and *Lactobacillus* spp. (Aas et al. 2008). Confirming the results of the previous study, Lima et al. (2011), using a checkerboard DNA-DNA hybridisation assay on caries samples in secondary dentition also detected association of *Atopobium* genomospecies C1, as well as *P. acidifaciens* with caries. This study, too, tried to differentiate microbiota from the different layers of the caries lesion and, while no significant differences were observed in the three dentine caries layers, the most prevalent taxa at the deepest layer were *F. nucleatum,* two *Lactobacillus* species as well as the G+C-rich species *Atopobium* genomospecies C1 (Lima et al. 2011).

A few studies have concentrated specifically on the detection of *Bifidobacteriaceae* (also members of the high G+C phylum *Actinobacteria*) in caries lesions using culture analysis, and Modesto et al. (2006) found the prevalence of *Bifidobacteriaceae* in both dental caries and dental plaque to be around 70 %. All isolates of *Scardovia inopinata* and most of *Parascardovia denticolens* produced acids from high-molecular-weight glucan, supporting the hypothesis *Bifidobacteriaceae* could play a role in the promotion of the caries process. Mantzourani et al. (2009) made contrasting observations in that they did not detect bifidobacteria in a culture based study from plaque in caries-free children and adults, which might be explained by differing culture techniques used in the two studies. Nonetheless, bifidobacteria were isolated from 13 of 15 caries lesions in adults and 16 out of 24 caries lesions in children, and it was argued that, since bifidobacteria showed pH lowering capacities comparable to *S. mutans* if the medium was glucose (Haukioja et al. 2008), bifidobacteria may have the potential to proliferate in caries lesions due to a favourable environment (Mantzourani et al. 2009). However, since bifidobacteria are found mainly in active lesions (Becker et al. 2002; Mantzourani et al. 2009) it seems likely they are involved in lesion progression rather than initiation. In fact, Beighton et al. (2010) argued bifidobacteria might be regarded as caries-associated organisms and their role in the caries process as well as markers of caries risk require further investigation.

As can be seen from all of these studies, several species have been associated with caries lesion development and/or progression. However, the findings of these recent studies have demonstrated that there is more to discover until a complete scheme can be developed describing which species have the potential to initiate or advance lesion formation and which species co-localise or form symbiotic networks in these processes. Moreover, it may be important not only to know which species are present (and their order of appearance), but also to characterise mechanisms of lesion initiation and/or progression. Findings of recent years suggest species belonging to the phylum *Actinobacteria* may have a significant role to play in caries lesion progression from the very beginning up until the caries lesion can be classified as advanced. In addition, observations have been made that microbiotic profiles of patients can be categorized in groups, potentially influencing the kind of preventive or curative treatment that is needed.

1.10 Molecular microbial ecology methodologies

1.10.1 Unculturability of bacterial species

As mentioned above, traditionally, bacteria were grown on solid media but it became clear not all species could be detected using culture analysis. 'The great plate count anomaly' is a well-known phenomenon where far lower counts of microorganisms were obtained when plate counts (i. e. colony forming units growing on solid media) were compared to those obtained from microscopy (Staley et al. 1985; Amann et al. 1995). Bacteria only grow when their nutritional and metabolic needs are being met and it is believed that for the majority of species the optimal culture conditions have not yet been found (Wade 2002; Vartoukian et al. 2010; Marsh et al. 2011). Indeed, it is known that some bacteria need other species for metabolic functions and thus cannot be grown in pure culture unless these specific metabolic needs are met (Wyss 1989; Vartoukian et al. 2010). It has been estimated that less than 2 % of bacteria on earth can be grown in culture, while in the oral cavity about 50 % of the microbiota are thought to be readily cultivable (Socransky et al. 1963; Wade 2002; Dewhirst et al. 2010).

1.10.2 Culture-independent methods

Given the predominance of unculturable bacteria, new methods to identify these bacteria were needed. A publication by Woese (1987) summarising fundamental work on bacterial evolution and the important role of the 16S rRNA gene as a taxonomic tool due to its role as an 'evolutionary clock', together with the invention of polymerase chain reaction (PCR) by Mullis et al. (1986) revolutionised the field of microbiology by allowing the identification of bacteria using molecular methodologies. The design and development of these molecular methods to phylogenetically characterise bacteria in any given habitat using the so-called housekeeping genes has helped overcome those issues of culturability and allowed identification of bacteria that previously proved refractory to detection.

The 16S rRNA gene is the target commonly chosen for phylogenetic identification using these molecular methodologies since it fulfils several significant criteria. Most importantly, it occurs in all organisms in the form of 16S rRNA in the case of *Bacteria* and *Archaea* and 18S rRNA in the case of eukaryotes (Woese et al. 1977). As Woese

(1987) described, its characteristic of showing clocklike behaviour makes it a useful phylogenetic chronometer, i. e. changes in its sequence occur randomly, while at the same time the molecule shows a high degree of functional constancy. Furthermore, the rates of change are in proportion to the spectrum of evolutionary distances measured and the molecule is large enough with highly conserved regions interspersed with variable ones to provide adequate amount of information (Woese 1987; Pace 1997). As a result, this gene represents a means of identifying organisms at different phylogenetic levels, down to species level.

However, the use of 16S targeted approaches also has drawbacks. The most important of these is probably the issue of assigning a sequence unambiguously to species level. As Janda et al. (2007) highlighted: there are no defined 'threshold values' (e.g. 98.5 % similarity) above which there is universal agreement of what constitutes definitive and conclusive identification to the rank of species. It is a well-known phenomenon that not all species can be assigned species-level identification using 16S rDNA sequencing. For example, Fox et al. (1992) found that *Bacillus globisporus* W25^T (T = type strain) and *Bacillus psychrophilus* (W16A^T, and W5) share a sequence identity of 99.5 %, while DNA-DNA hybridisation results prior to 16S rDNA sequencing established that these strains do not belong to the same species. Further examples of these so-called fuzzy species are *Neisseria* (Hanage et al. 2005), *Streptococcus* (Hanage et al. 2006) and *Actinomyces* (Henssge et al. 2009), amongst others. In these cases, species identification can be achieved by applying a multi-locus sequence typing (MLST) assay using an array of slow-evolving housekeeping genes (other than 16S rDNA) (Urwin et al. 2003; Hanage et al. 2006; Henssge et al. 2011).

The second component of the culture-independent analysis, polymerase chain reaction (**Error! Reference source not found.**), uses thermostable DNA polymerases (e.g. isolated from *Thermus aquaticus*) for the amplification of targeted DNA segments *in vitro*. For this, following denaturation of double stranded DNA, short oligonucleotide primers anneal to the single stranded target sequence and the polymerase subsequently lengthens the small double stranded segment in the elongation / amplification step. The repetition of these steps allows amplification of the target segment by 10^6-fold or more (Saiki et al. 1988; Wilson et al. 1990). If universal or broad range primers targeting conserved regions are used to amplify 16S rRNA genes from

environmental samples subsequent cloning is needed, before sequences can be identified using sequencing techniques.

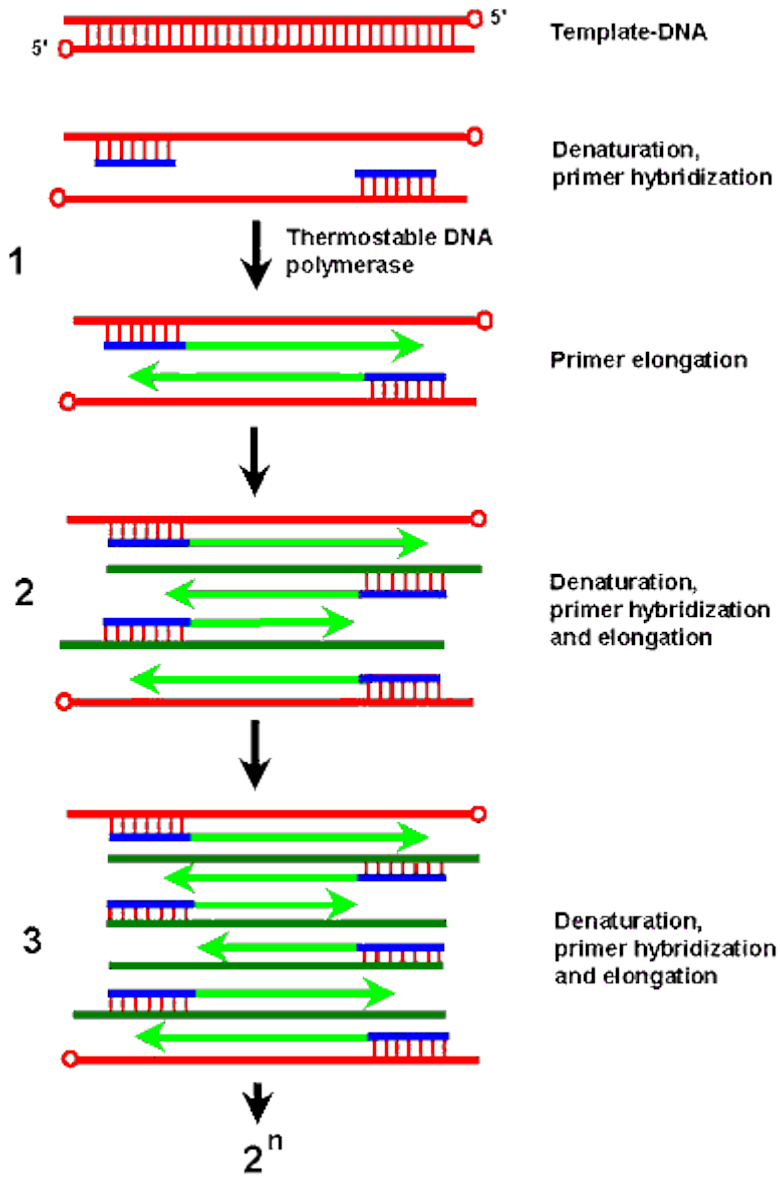


Figure 1.2: Schematic representation of PCR reaction (source: http://www.bioscience.org/1996/v1/e/reischl1/images/fig3.gif)

1.10.2.1 Applications of culture-independent methods

Though it is possible to directly sequence rRNAs (Stahl et al. 1985; Edwards et al. 1989) or to sequence cDNA libraries constructed from 16S mRNA (Ward et al. 1990), the strategy of 16S rDNA amplification, followed by subcloning into a vector and subsequent sequencing of the target DNA from these clones has become the method of choice for many research studies. Indeed, Sanger sequencing has been used in the

phylogenetic analysis of bacterial 16S rRNA genes to explore the biodiversity of habitats as varied as marine environments (Giovannoni et al. 1990; Devereux et al. 1994; Gray et al. 1996), hot springs (Papke et al. 2003), soil (Borneman et al. 1996), animal intestines (Whitford et al. 1998; Leser et al. 2002) as well as human intestines (Suau et al. 1999; Eckburg et al. 2005).

Many studies characterising the oral microbiota in health or disease have used the PCR / cloning / sequencing methodology (Aas et al. 2005; Aas et al. 2008) and have identified significant differences to what was previously discovered using culture techniques. For example, Choi *et al.* (1994) analysed the diversity of cultivable and uncultivable oral spirochaetes from a patient with severe destructive periodontitis and detected 20 novel *Treponema* 'species'. Similarly, Dymock *et al.* (1996) analysed the microbiota associated with dentoalveolar abscesses and found novel phylotypes representing uncultured species and also showed that species including *Fusobacterium nucleatum* and *Porphyromonas endodontalis* were underrepresented in culture. A further study investigated the bacterial diversity in human subgingival plaque from healthy subjects and subjects with refractory periodontitis, adult periodontitis, human immunodeficiency virus periodontitis, and acute necrotising ulcerative gingivitis (Paster et al. 2001). In this study the majority of clones (60 %) could be assigned to 132 known species, while 40 % of the clones represented novel phylotypes (Paster et al. 2001). Furthermore, periodontitis-associated pathogens such as *Porphyromonas gingivalis*, *Tannerella forsythia*, and *Treponema denticola* were identified from multiple subjects, but typically as a minor component of the plaque as seen in cultivable studies (Paster et al. 2001).

1.10.2.1.1 Bias caused by high G+C DNA

Numerous studies, described in section 1.9.4 (Microbiology of Caries), investigating the microbiota of dentine caries have used this approach for the identification of microbiota (Byun et al. 2004; Munson et al. 2004; Nadkarni et al. 2004; Chhour et al. 2005). However, studies assessing the microbiota associated with dental caries (Munson et al. 2004), endodontic infections (Munson et al. 2002) and subgingival plaque (de Lillo et al. 2006), which all included a direct comparison of the sequencing data to culture, reported the under-estimation of the proportion of species with a high

G+C content, such as members of the phylum *Actinobacteria*. Thus, there are problems associated with this methodology.

A likely explanation for this phenomenon is the low efficiency of high G+C DNA strand separation and/or the occurrence of secondary structures that causes the *Taq* polymerase to pause or prematurely terminate elongation (McConlogue et al. 1988; Wilson et al. 1990). As a consequence, DNA with a lower G+C content that readily denatures under standard PCR conditions is amplified preferentially when samples consist of both low G+C and high G+C DNA (Varadaraj et al. 1994; Wagner et al. 1994; Mytelka et al. 1996; Hill et al. 2006). Dutton et al. (1993) published an optimised protocol consisting of a high temperature denaturation, a combined annealing/elongation step, the use of very heat stable thermal polymerases and oligonucleotides with a Wallace temperature[1] of 80 °C or 120 °C. This protocol was suitable for the amplification of DNA with a G+C content of up to 75 % as it circumvented problems associated with secondary structures in the template. It should be noted, however, that for segments with G+C content significantly greater than 75 % it may be necessary to increase temperatures for denaturation and annealing/elongation or to add chemical denaturants (Dutton et al. 1993).

One such chemical denaturant was found to be betaine (Henke et al. 1997). One of the advantages of this molecule is that it works with conventional *Taq* polymerases as well as with PCR assays designed for hot start PCR or long and accurate PCR (Henke et al. 1997). In addition, it was also found that the addition of betaine reduces polymerase pauses, thereby increasing efficiency and probably processivity. A similar approach was reported by Baskaran et al. (1996) who found a uniform amplification of heterogenous DNA templates (44 – 80 % G+C) was only achieved when a combination of betaine and DMSO was used. A more recent study by Sahdev et al. (2007) reported a successful combination of these strategies, incorporating primer modifications, DMSO-betaine combinations and high denaturing temperature conditions. Other recent studies have also reported the application of trehalose as a potent PCR enhancer for GC-rich

[1] Wallace temperature: a formula to calculate the melting temperature T_d of oligonucleotides at 0.9M NaCl, based on the number of occurrences of each nucleotide. $T_d = 2°C(A+T) + 4°C(G+C)$

templates (Spiess et al. 2004; Horakova et al. 2011). It is believed trehalose lowers the template melting temperature and eliminates secondary structures as well as thermostabilising the *Taq* polymerase (Spiess et al. 2004). Comparable to the case of DMSO/betaine mix, a recent study found a supplement of 1 M 1,2-propanediol in conjunction with 0.2 M trehalose greatly enhanced amplification results of G+C-rich templates; moreover, it was found to efficiently neutralise PCR inhibitors present in blood samples (Horakova et al. 2011).

A slightly different approach was proposed by Frey et al. (2008) who reported a protocol for extremely G+C-rich (up to 84 %) templates based on the so-called 'touchdown' PCR invented by Don et al. (1991). The new 'slowdown' PCR optimises primer annealing by decreasing the annealing temperature every third cycle by +1 °C at an overall reduced heating ramp rate of 2.5 °C s^{-1} combined with a slow cooling rate of 1.5 °C s^{-1} to reach the annealing temperature. The protocol furthermore relies on the inclusion of the additive 7-deaza-2′-deoxyguanosine, a dGTP analogue, to the PCR mixture. This was one of the first additives to be reported as improving PCR amplification of templates with a high G+C content (McConlogue et al. 1988). Frey et al. (2008) tested the protocol on the *BRAF* gene, which had escaped amplification and sequencing despite the use of five different primer pairs and only using the above protocol was a single specific band of the correct size generated, whereas standard or 'touchdown' PCR failed. Another study has since been published reporting the modification of the 'touchdown' PCR and addition of an additive (betaine) for the enhanced amplification of GC-rich templates (Pratyush et al. 2012).

Yet another strategy of tackling this problem is the use of alternative polymerases, for which the manufacturers' have made claims proposing their suitability for use with high G+C templates. The Phusion™ High-Fidelity DNA Polymerase (New England Biolabs) product uses a novel *Pyrococcus*-like enzyme with a processivity-enhancing domain together with GC buffer which includes additives, not listed by the manufacturer, that enhance the amplification of G+C-rich regions. Herculase® II Fusion polymerase (Stratagene) also amplifies targets containing GC-rich regions of up to 84%, since its double stranded binding domain keeps it bound to the template and greatly enhances rates of processivity. PrimeSTAR HS, a high fidelity DNA polymerase, promising high accuracy due to a strong exonuclease activity in conjunction with an

optimised GC Buffer (Takara) was developed for high-fidelity amplification of GC-rich (greater than or equal to 75 %) templates. The AccuPrime™ GC-Rich DNA Polymerase (Life Technologies), a thermostable polymerase boasting a five-fold better processivity than *Taq* polymerase, comes with a choice of buffers that contain thermostable proteins that enhance primer-template hybridisation during PCR, increasing the specificity of the reaction.

A further polymerase, TopoTaq (Fidelitysystems), was developed after it was found that the DNA-binding helix–hairpin–helix (HhH) motifs in DNA topoisomerase V (Topo V) have a great impact on its activity and processivity at high salt concentrations (Pavlov et al. 2002). Subsequently hybrid proteins consisting of the Stoffel fragment of *Taq* DNA polymerase or *Pfu* DNA polymerase and different TopoV's HhH subdomains fused with either the NH_2- or the COOH terminus of polymerases were designed. It is claimed that the resultant TopoTaq polymerase dramatically advances the key properties of the *Taq* DNA polymerase such as processivity, thermostability and specificity, and that it is resistant to common inhibitors of PCR, including DNA intercalating dyes, organic solvents, and biological fluids. The performance of TopoTaq is further enhanced by addition of a hyperstable *Methanopyrus* DNA topoisomerase that facilitates DNA strand separation.

1.10.2.2 Next generation sequencing

Until recently, sequencing reactions were performed using the so-called Sanger method (Figure 1.3). In addition to the normal deoxynucleotides (dNTPs), this method uses dideoxynucleotides (ddNTPs), which have a hydrogen group on the 3’ carbon instead of a hydroxyl group. When ddNTPs are included in a DNA segment no further NTPs can be added as no phosphodiester bond can be formed and the DNA chain elongation is consequently terminated. In automated reactions, all four ddNTPs are labelled with different coloured dyes and subsequently run through a capillary, separating the DNA fragments by length. A laser then reads each fragment emerging from the capillary and identifies the labelled nucleotide according to the wavelength at which it fluoresces, resulting in a graphic output in the form of a chromatogram. The disadvantage of this method is that it is time consuming and labour intensive. Only relatively few samples can be included in studies and it has become clear that the full

species richness of a sample cannot be appreciated using this method. In other words, the rare biosphere remains undetected (Sogin et al. 2006; Pedros-Alio 2012).

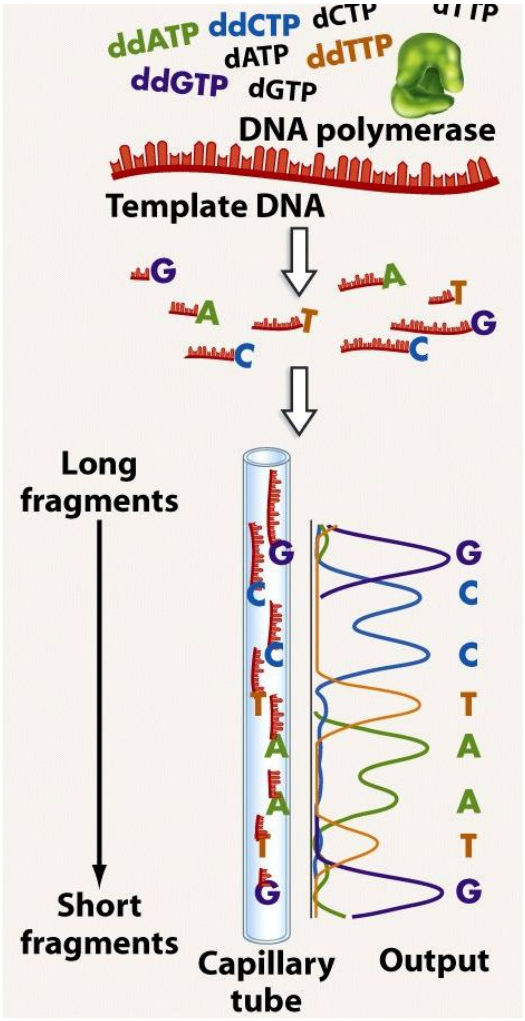


Figure 1.3 Schematic representation of Sanger sequencing (adapted from http://www.uic.edu/classes/bios/bios100/lectures/sanger03.jpg)

As a result, new culture-independent methods for phylogenetic analysis of the microbiota have been developed in recent years. One of these 'next-generation' sequencing techniques is known as the Illumina platform. This system uses a sequencing by-synthesis approach (Mardis 2008). DNA fragments that are attached to primers on a slide are amplified *in situ* before the so-called reversible terminator bases (RT-bases) are added, incorporating a fluorescent label. Once a nucleotide has been incorporated, all other nucleotides are washed away and a camera records which dye, and therefore which nucleotide, has been added. Before the next nucleotide can be

added, the dye and terminal 3' blocker are removed in a chemical reaction. The Illumina platform has been used to study the oral microbiota (Lazarevic et al. 2009), though because of the short read lengths achieved with this method, another high-throughput sequencing method resulting in substantially longer reads has become the method of choice to analyse oral microbiota. This method is known as pyrosequencing, or 454 sequencing.

Pyrosequencing uses emulsion PCR, in which DNA is amplified in oil:water micelles containing all PCR reagents and a primer-coated bead (Figure 1.4). Following emulsion PCR, the micelles are broken and a single cloned amplified DNA bead is deposited per well on a picotiter plate (PTP). Beads containing pyrosequencing reagents are subsequently added and since the incorporation of a nucleotide is accompanied by a release of pyrophosphate (PPi) in equimolar amounts to that of the incorporated nucleotide, ultimately resulting in the generation of visible light in a downstream chemiluminescent reaction, a charge coupled device (CCD) records a peak in the raw data output. Following each potential insertion reaction, an enzyme degrades any remaining unincorporated nucleotides and thereof resulting ATP before a new nucleotide is added. The intensity of the light signal, and therefore the height of each recorded peak is proportional to the number of incorporated nucleotides, from which the exact DNA sequence can be deduced (Mardis 2008).

One of the first studies applying this method to study the oral microbiota were Keijser et al. (2008). Approximately 200 000 amplicons were sequenced and found to represent 5 669 phylotypes in saliva and just over 10 000 in dental plaque samples, at a sequence identity cut-off of 97 %. Chao1 values, a measure of minimum species richness in a community based on the OTU definition, were estimated to be 12 611 and 26 204 for saliva and plaque respectively, far higher than previously reported (Keijser et al. 2008). It has since been recognised that homopolymers (repeated single bases) lead to erroneous base calling (i. e. assigning signal to a sequence representing the nucleotides, or bases), and together with the problem of formation of chimeric sequences during PCR, this can result in the overestimation of species richness.

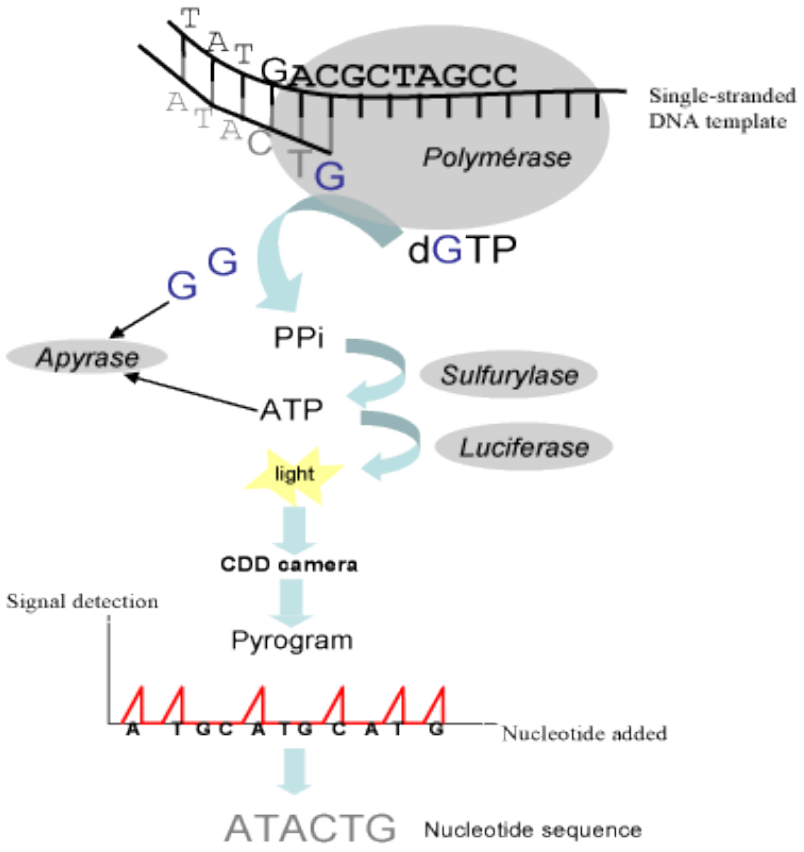


Figure 1.4: Schematic representation of 454 sequencing. (Source (Armougom et al. 2009)

Quince et al. (2009) addressed this problem and presented an algorithm, PyroNoise, that allowed accurate calculation of operational taxonomic units. Since the above-mentioned study, which used 100 nt long DNA sequences, the technology has been improved and newer versions of the 454 platform can generate up to 500 bp long sequences, making it more comparable to Sanger sequencing read lengths and therefore result in better phylogenetic resolution. A further advantage of pyrosequencing is the avoidance of bacterial cloning, which means sequences not usually detected due to cloning bias are more likely to be represented, leading yet again to a truer representation of the sampled habitat (Mardis 2008).

One of the main challenges of high throughput sequencing is the amount of data generated and its subsequent analysis. This was highlighted by Li et al. (2012) who examined the microbial diversity across body habitats (oral, skin, distal gut, and vaginal body regions) from over 200 healthy individuals and found that the variation of diversity within low abundance taxa across habitats and individuals was not quantified

sufficiently with standard ecological diversity indices. It was concluded that analysis tools specifically for next-generation sequencing data are needed, such as the proposed Tail statistic, which may help account for low abundant taxa (Li et al. 2012). On a similar note Diaz et al. (2012) emphasised the importance of knowing the limitations of next generation sequencing and its data analysis, since it was found that the commonly used Good's coverage estimator, which is based on sequences that occur only once, vastly underestimated coverage. The handling of singletons in these data sets remains a point for debate and inclusion or deletion of singletons require careful consideration (Reeder et al. 2009; Schloss et al. 2011; Diaz et al. 2012).

1.10.2.3 Fluorescence in situ hybridisation for the detection of oral bacteria

Knowing which species are present is of course vital to assess any association with health or disease and not just in the oral cavity. However, simply knowing of the existence of a species in a disease state or the association of two or more species with each other does not prove they are colonising the same area, let alone confirm an interaction. For this, a different approach is required. Simultaneous visualisation, identification (to species or higher taxon levels) and localisation of individual microbial cells is possible using fluorescent *in situ* hybridisation (FISH). For the application of this technique an oligonucleotide probe of chosen specificity with a fluorescent label is designed that binds to its complementary target sequence within the intact, fixed cell (Amann et al. 1995; Moter et al. 2000). The 16S rRNA gene is the target of choice in most microbiology studies due to its aforementioned conserved and variable regions, as well as its high copy number. This allows detection of not yet cultivated phylotypes, thereby making it a culture-independent method. Visualisation of fluorescence can be achieved using an epifluorescence microscope or, as is now standard procedure, a confocal laser scanning microscope (CLSM) (Amann et al. 1995; Moter et al. 2000).

Since probes only target intact and living cells and can penetrate target cells in undisturbed biofilms, this technique was soon applied to examine the three-dimensional architecture of dental plaque. Netuschil et al. (1998) were the first to publish a study in which visualisation of the three-dimensional topography of the biofilm using confocal microscopy was combined with fluorescence staining of the dental plaque flora for assessment of its vitality.

FISH has now been used extensively, especially for the examination of early supragingival plaque formation. Applications of this technique range from the quantification of cells of initial bacterial colonisation on enamel (Hannig et al. 2007) to the visualisation of the early biofilm architecture during the first few hours (Diaz et al. 2006; Al-Ahmad et al. 2009; Jung et al. 2010), as well as the characterisation of early biofilm development over several days (Al-Ahmad et al. 2007; Dige et al. 2007; Quevedo et al. 2011). Few studies, however, have focused on very mature biofilm architecture (Zijnge et al. 2010) or used this technique to study the prevalence and localisation of bacteria in carious dentine (Banerjee et al. 2002).

Various analysis methods to study the species growing as biofilms in the oral cavity have been designed and applied. For example, *in vitro* models, such as growth of biofilms on hydroxyapatite discs (Thurnheer et al. 2004) or in saliva-conditioned flow cells (Foster et al. 2004) have been developed. In another approach samples were excavated for the detection and quantification of bacteria in carious dentine (Banerjee et al. 2002). However, these methods have the disadvantage of not being able to truly mimic the conditions and occurrence of species found in the oral cavity and the excavation of material will likely disturb the natural localisation of cells. Consequently *in situ* analysis methods have been developed in which subjects wear slabs made of human (Diaz et al. 2006) or, more commonly, bovine enamel (Hannig et al. 2007) or dentine (Jung et al. 2010) or glass (Dige et al. 2007) intraorally. This approach ensures the surfaces to be colonised are subjected to all prevailing environmental conditions and species in the mouth, while the test surface can be removed and analysed at the end of the test period without disturbing the natural biofilm architecture.

A further advantage of FISH is that it allows cells to be targeted with universal and specific probes at specified taxonomic levels. Accordingly, this has made possible the use of species- or genus-specific probes in conjunction with a universal probe to assess proportions of the targeted cells in comparison to total cells (Hannig et al. 2007; Al-Ahmad et al. 2009; Jung et al. 2010). This, of course, is in addition to the information gained concerning (co-)localisation of bacteria in the early (Diaz et al. 2006; Dige et al. 2007; Al-Ahmad et al. 2009) or maturing biofilm (Guggenheim et al. 2001; Zijnge et al. 2010), which may help in the development of clinical treatments that disrupt

coaggregation or other forms of bacterial-interdependencies, thereby halting or even preventing disease initiation or progression.

1.11 Aims

To date, the progression of which bacterial species which initiate and drive progression of dental caries, and in which manner, is not fully understood. Since the development of new measures to prevent and/or treat dental caries relies on the complete understanding of this disease and its associated species, continued efforts are required to characterise the caries microbiota and their interactions with each other and/or their environment.

The aims of this study were:

i) to assess the source of bias in PCR amplifications of environmental samples with regards to the underrepresentation of species with a high G+C DNA content and to design and evaluate a method to overcome those biases.
ii) to characterise the microbiota of dental caries samples using a combination of culture and PCR / cloning / sequencing approaches developed above.
iii) to revisit the microbiology of the above caries lesion samples using 454 pyrosequencing and to subsequently evaluate whether this method improved overall detection of species, as well as detection of high G+C species in particular, compared to Sanger based sequencing.
iv) to design and validate oligonucleotide probes of chosen bacterial taxa for fluorescent *in situ* hybridisation (FISH) to study the prevalence and location of those species in the caries lesion.

Chapter 2

Investigation of sources of bias in molecular characterisation of oral bacterial communities

2 Chapter 2

2.1 Introduction

It has been observed in a number of culture-independent microbial ecology studies that molecular analysis has under-estimated the proportion of species with a high G+C content, such as members of the phylum *Actinobacteria* (Munson et al. 2002; Munson et al. 2004; de Lillo et al. 2006). One possible reason for this is that *Taq* polymerase is prone to pausing or prematurely terminating elongation when encountering G+C-rich regions of template. This can be due to the low efficiency of high G+C DNA strand separation and/or the occurrence of secondary structures (McConlogue et al. 1988; Wilson et al. 1990). Consequently, *Taq* polymerase preferentially amplifies DNA with a lower G+C content, that readily separates under standard PCR conditions when samples consist of both low G+C and high G+C DNA (Varadaraj et al. 1994; Wagner et al. 1994; Mytelka et al. 1996; Hill et al. 2006).

Additives have been used to improve amplification of high G+C regions. Tetramethylammonium chloride (TMAC), for example, was used by Hung et. al. (1990) to improve stringency in primer hybridisation and thereby to eliminate non-specific amplification. To facilitate DNA strand separation and therefore improve the amplification of DNA templates with a high G+C content, betaine was added due to its isostabilising effects (Rees et al. 1993). DMSO, 7-deaza-2-deoxyguanosine and formamide have been used because they disrupt base pairing (e. g. disrupt hydrogen bonding) (McConlogue et al. 1988; Bookstein et al. 1990; Sarkar et al. 1990; Baskaran et al. 1996).

Alternative polymerases are also available, for which the manufacturers' have made claims proposing their suitability for use with high G+C templates. The Phusion™ High-Fidelity DNA Polymerase (New England Biolabs) product uses a novel *Pyrococcus*-like enzyme with a processivity-enhancing domain together with GC buffer which includes additives, not listed by the manufacturer, that enhance the amplification of G+C-rich regions. Herculase® II Fusion polymerase (Stratagene) also amplifies targets containing GC rich regions of up to 84%, since its double stranded binding domain keeps it bound to the template and greatly enhances rates of processivity.

Consequently, the widely used *Taq* polymerase was used alongside the Phusion and Herculase polymerases in this chapter to compare their ability to amplify DNA extracted from individual bacterial species with low and high G+C DNA and mixtures thereof.

Due to the nature of the variant composition of the bacterial cell wall the choice of DNA extraction protocol can have an impact on the subsequent PCR amplification and therefore detection of bacterial species from environmental samples, as has been observed in various studies (Rantakokko-Jalava et al. 2002; Purswani et al. 2011). Therefore, five DNA extraction protocols have been evaluated regarding their potential of skewing representation of high G+C organisms in molecular analysis. Among those was the widely used phenol-chloroform extraction method adapted from a protocol by Marmur (1961), the other four comprising commercially available kits.

A further component potentially introducing a bias to the molecular analysis approach is the choice of primers. The universal primer set 27F CM / 1492R (Lane 1991), specific for the domain *Bacteria*, has been used in numerous molecular studies. However, it has been demonstrated that it is not truly universal for this domain as the 16S rRNA genes of many species are not amplified (Marchesi et al. 1998). A new primer set, 63F / 1387R was designed and validated successfully and was demonstrated to have a broader range than 27F CM / 1492R (Marchesi et al. 1998).

Frank *et al.* (2008) re-evaluated the primer pair 27F CM / 1492R and their results support the use of 1492R as a universal bacterial primer. Whilst the non-degenerate primer 27F-CC (5´-AGAGTTTGATCCTGGCTCAG-3´) was found to match the binding site sequence of most bacteria listed in the Ribosomal Database Project (RDP) and from Sargasso Sea data (Giovannoni et al. 1990), mismatches with many bacteria were observed, particularly of the Sargasso Sea rRNA genes (Frank et al. 2008). Another commonly used primer, 27F-CM (5´-AGAGTTTGATCMTGGCTCAG-3´), includes two "wobble" bases that match the most common binding site variants, covering most of the bacterial phyla (Frank et al. 2008). Nonetheless, further binding site sequence variants, such as those found in the phylogenetic groups of *Actinobacteria*, some *Proteobacteria*, *Campylobacteriales* and *Sphingomonadales*, were only accommodated by the primer 27F-YM (5´-AGAGTTTGATYMTGGCTCAG-3´)(Frank et al. 2008). An

alignment specific to oral bacteria revealed a new primer candidate which was subsequently successfully validated and incorporated in this study alongside the other primers.

2.2 Aims

The aims of the work described in this chapter are:

i) To compare of the ability of *Taq* and *Pfu* based polymerases to amplify 16S rRNA genes from Gram positive species with DNA of low (*S. sanguinis*) and high (*A. naeslundii*) G+C content.

ii) To determine the influence of five different DNA extraction methods on the amplification by PCR.

iii) To compare the influence of various primer pairs to amplify 16S rRNA genes from species with DNA of low and high G+C content and the *S. sanguinis / A. naeslundii* mixture.

2.3 Materials and methods

2.3.1 Bacterial strains and culture

The bacterial strains used are shown in Table 2.1. The strains were obtained from the departmental culture collection and cultured on Blood Agar (BA, Blood Agar base No.2, Bioconnections, Knypersley, UK), supplemented with 5 % sterile defibrinated horse blood (TCS Biosciences Ltd, Buckingham, UK), under anaerobic conditions at 37 °C in an anaerobic workstation (Don Whitley Scientific Ltd, Shipley, UK) with an atmosphere of 80 % nitrogen, 10 % hydrogen and 10 % carbon dioxide. For quantitation and DNA extraction, strains were grown under anaerobic conditions in Brain Heart Infusion broth (BHI, LabM, Heywood, UK). All media were prepared by autoclaving at 121 °C for 15 min.

Table 2.1: Bacterial strains

Species	Strain no.	G+C content (mol%)
Actinomyces naeslundii	NCTC 10301	66
Streptococcus sanguinis	ATCC 10556^T	46
Atopobium parvulum	ATCC 33793	45.7
Bifidobacterium dentium	NCTC 11816	61.2
Olsenella uli	ATCC 49627	63.5

2.3.2 Calibration curves of optical density plotted against viable counts for *A. naeslundii* and *S. sanguinis*.

After incubation, cultures of *A. naeslundii* and *S. sanguinis* were diluted ten-fold and the optical density (OD) measured at a wavelength of 550 nm. Aliquots (50 µl and 100 µl) of the diluted broth cultures were spread on to BA, incubated under anaerobic conditions at 37 °C over night in the case of *S. sanguinis* and for 48 h in the case of. *A naeslundii*, and viable counts registered. The OD_{550nm} values, measured in duplicate for each species, were plotted against the corresponding viable counts to plot a calibration curve. Thus, suspensions of known colony forming units (cfu) / ml could be prepared by adjusting the turbidity of the broth cultures without having to perform viable counts.

2.3.3 DNA extraction

DNA was extracted in five different ways: a conventional method based on the Marmur method using proteinase K and phenol-chloroform (Munson et al. 2004; Sakata et al. 2006), as well as by means of four different extraction kits according to the manufacturer's instructions. These comprised of the GenEluteTM bacterial genomic DNA extraction kit (Sigma Aldrich, Gillingham, UK)

2.3.4 Phenol-chloroform DNA extraction

One ml of bacterial suspension grown over night was centrifuged for 5 min at 13 000 g in a Biofuge fresco (Heraeus) and harvested cells resuspended in 400 µl 1x TE buffer (10 mM Tris-HCl containing 1 mM EDTA, pH 8.0, Sigma Aldrich). The cell suspension was treated with 25 µl 10 mg/ml lysozyme (Sigma Aldrich) for 30 min at 37 °C, after which 25 µl 10 mg/ml Proteinase K (Sigma Aldrich) and 50 µl 20 % Sarkosyl solution (N-lauroylsacrosine, Sigma Aldrich) were added and the tube briefly vortex mixed (Jencons-PLS VX1000, Jencons Scientific Ltd, Bath, UK). The suspension was incubated at 60 °C for 1 h, inverting the tube once after 30 min, and was then cooled on ice. An equal volume of 25:24:1 phenol:chloroform:isoamyl alcohol (Sigma-Aldrich) was added and the suspension mixed by shaking. Following centrifugation at 13 000 g for 5 min the supernatant was aspirated into a fresh Eppendorf tube and the phenol chloroform treatment repeated. One ml of ice cold 99.6 % ethanol was added to the extract and incubated at -20 °C for 15 min. After centrifugation at 13 000 g for 5 min the supernatant was carefully discarded and 1 ml ice-cold 70 % ethanol added to the pellet. The tube was inverted carefully several times and centrifuged again at 13 000 g for 5 min. The supernatant was carefully pipetted off and the pellet left to dry at room temperature. Finally, the pellet was resuspended in 100 µl 1x TE buffer.

2.3.5 GenElute™ DNA extraction

Cells were harvested by centrifuging 1.5 ml of an overnight bacterial broth culture at 13 000 g for 2 min. The culture medium was discarded, the pellet resuspended in 200 µl 45 mg/ml lysozyme solution (Sigma Aldrich) and incubated at 37 °C for 30 min. Twenty µl RNase A solution was added and incubated at room temperature for 2 min, after which 20 µl Proteinase K and 200 µl Lysis solution C were added. The suspension was vortex mixed and incubated at 55 °C for 10 min in a Dri-Block (DB-3, Techne, Stone, UK). Meanwhile columns were prepared by adding 500 µl Column Preparation

Solution to each preassembled GenElute™ Miniprep Binding Column and collection tube and centrifuging at 13 000 g for 1 min. The eluate was discarded. Following incubation at 55 °C, 200 µl 99.6 % ethanol was added to the lysate and mixed thoroughly by vortex mixing for about 10 sec. The complete lysate was transferred to the binding column using a wide bore pipette tip to minimize shearing the DNA and centrifuged at 13 000 g for 1 min. The column was then placed in a fresh 2 ml collection tube and 500 µl Wash Solution 1 added. After centrifuging for a further minute at 13 000 g the column was again placed in a fresh 2 ml collection tube and 500 µl of Wash Solution concentrate diluted with the appropriate amount of ethanol added. The column was centrifuged once for 3 min at 13 000 g, the collection tube emptied and the column spun a further minute at 13 000 g to dry it thoroughly. The column was placed in a fresh collection tube and 200 µl of the Elution Solution were pipetted directly onto the centre of the column. After a 5 min incubation at room temperature to increase the elution efficiency as suggested by the manufacturer, the column was centrifuged at 13 000 g for 1 min. The elution process was repeated to increase yield and a 50 µl aliquot was stored at 4 °C for short term usage, while the remainder was stored at -70 °C.

2.3.6 microLysis DNA extraction

Cells were harvested by centrifuging 200 µl of an overnight bacterial broth culture at 13 000 g for 2 min. The culture medium was discarded and the pellet resuspended in 100 µl 0.1x TE buffer. A 3 µl aliquot of cells of undetermined concentration or 10 µl of cells at a concentration of 1, 2 or 4 x 10^7 cfu/ml were mixed with 17 µl and 20 µl microLYSIS solution (Microzone, Haywards Heath, UK) , respectively. The PCR tube with the cell solution was placed in a thermal cycler and the cycling profile consisted of step 1) 65 °C for 5 min, 2) 96 °C for 2 min, 3) 65 °C for 4 min, 4) 96 °C for 1 min, 5) 65 °C for 1 min, 6) 96 °C for 30 sec, 7) 99 °C for 10 min and the final step 8) hold at 20 °C. The extracted DNA was used for PCR applications straight away, but could also be stored at -20 °C.

2.3.7 DNA extraction using the FastDNA kit and FastPrep24 homogenizer

A. naeslundii and *S. sanguinis* cells were harvested by centrifuging 1 ml of an overnight bacterial broth culture at 13 000 g for 2 min. The culture medium was discarded and the pellet resuspended in 200 µl 0.1x TE buffer. The sample was added to matrix B (MP

Biomedicals, Cambridge, UK) and 1 ml of solution CLS-TC added. Samples were homogenised at 6 m/sec 3 x 30 and 40 sec, respectively. Vials containing homogenised cell culture were kept on ice between runs for 2 min. Cells were then pelleted for 10 min at 13 000g, the supernatant transferred to a fresh tube and 20 µl RNase added to each tube and incubated for 2 min at room temperature. Following this the DNA was extracted using the phenol-chloroform treatment three times as described in section 2.3.4. Finally, the DNA was resuspended in 100 µl 0.1x TE buffer.

2.3.8 MOBIO UltraClean Microbial DNA extraction

Broth-culture aliquots (1.8 ml) were centrifuged at 10 000 g for 5 min. The supernatant was discarded and the sample centrifuged a further 2 min at 10 000 g and supernatant carefully removed with a pipette. The cell pellet was resuspended in 300 µl MicroBead solution and gently vortex mixed. The solution was transferred to the MicroBead tube (MoBio, Carlsbad, CA, USA) and 50 µl solution MD1 added. Samples were heated at 65 °C for 10 min, vortex mixed for 10 min at maximum speed and tubes finally centrifuged for 30 sec at 10 000 g. The supernatant was transferred to a fresh tube and 100 µl solution MD2 added. Samples were vortex mixed for 5 sec and incubated at 4 °C for 5 min, before they were centrifuged for 1 min at 10 000 g. The supernatant was again transferred to a fresh tube and 900 µl solution MD3 added and vortex mixed for 5 sec. About 650 µl of the solution were loaded into a SpinFilter and centrifuged for 30 sec at 10 000 g. The flow through was discarded, the procedure repeated with the remaining supernatant and 300 µl solution MD4 added before samples were centrifuged for 30 sec at 10 000 g. The flow through was discarded and tubes centrifuged for 1 min at 10 000 g. The SpinFilter was carefully placed into a fresh tube and 50 µl of solution MD5 added to the centre of the filter. Samples were centrifuged a final time for 30 sec at 10 000 g and the SpinFilter discarded.

2.3.9 Separation and visualisation of DNA and PCR products

DNA samples, PCR products and molecular weight markers (100 bp DNA ladder, NEB, Hitchin, UK) were subjected to electrophoresis in 1 % agarose gels prepared with molecular biology grade agarose (Bioline, London, UK) dissolved in 0.5x Tris-Borate EDTA (TBE) buffer (Sigma-Aldrich) containing 0.5 µg ml^{-1} ethidium bromide (Amresco Inc., Solon, OH, USA) or GelRed (Biotium, Hayward, CA, USA). Electrophoresis was carried out at 105 V for 45 min. Ethidium bromide/GelRed stained DNA bands were

observed under UV light (302 nm) using a 3UV Transilluminator (UVP) and photographed.

DNA was quantified by visual comparison with known amounts of lambda phage DNA (NEB).

2.3.10 PCR amplification of 16S rRNA genes

PCR reactions were performed in a PxE 0.2 Thermal Cycler (Thermo Scientific, Epsom, UK) or a TC-412 thermal cycler (Techne).

2.3.10.1 Thermoprime Taq PCR

Reactions were prepared containing 23 µl Thermoprime *Taq* polymerase master mix (Thermo Scientific), 1 µl of template and 0.5 µl of each primer (10 µM, Table 2.2). Initial denaturation was at 95 °C for 5 min, followed by 30 cycles of denaturation at 95 °C for 45 sec, annealing at 56 °C for 45 sec and extension at 72 °C for 90 sec.

2.3.10.2 Herculase II fusion PCR

Reactions were prepared containing 10 µl 5× Herculase buffer, 1.25 µl 10 mM dNTPs, 1 µl Herculase fusion II polymerase, 1.25 µl of each primer (10 µM, Table 2.2), 1 µl of template and 34.25 µl sterile water. Initial denaturation was at 98 °C for 2 min, followed by 30 cycles of denaturation at 98 °C for 20 sec, annealing at 56 °C for 20 sec and extension at 72 °C for 30 sec.

2.3.10.3 Phusion High fidelity Hot Start PCR

Reactions were prepared containing 4 µl 5× Phusion buffer GC, 0.4 µl 10 mM dNTPs, 0.2 µl Phusion HF polymerase (0.4 U, Finnzymes), 0.5 µl of each primer (10 µM, Table 2.2), 1 µl of template and 13.4 µl sterile water. Initial denaturation was at 98 °C for 30 sec, followed by 30 cycles of denaturation at 98 °C for 10 sec, annealing at 56 °C for 30 sec and extension at 72 °C for 45 sec.

2.3.10.4 PCR primers

Table 2.2: Primers (IUPAC notation of degenerate bases: M = A or C; Y = C or T; R = A or G; W = A or T)

Primer	Sequence	
27F CM	5´-AGAGTTTGATCMTGGCTCAG-3´	(Lane 1991)
27F YM	5´-AGAGTTTGATYMTGGCTCAG-3´	(Frank et al. 2008)
Bif27F	5´-AGGGTTCGATTCTGGCTCAG-3´	This study

39F	5´-ATCMTGGCTCAGRWYGAACGC-3´	This study
61F	5´-CAGGCCTAACACATGCAAG-3´	This study
63F	5´-CAGGCCTAACACATGCAAGTC-3´	(Marchesi et al. 1998)
1387R	5´-GGGCGGWGTGTACAAGGC-3´	(Marchesi et al. 1998)
1492R	5´-TACGGYTACCTTGTTACGACTT-3´	(Lane 1991)
M13 Forward (-20)	5´-GTAAAACGACGGCCAG-3´	Invitrogen
M13 Reverse	5´CAGGAAACAGCTATGAC-3´	Invitrogen

2.3.11 Library creation

2.3.11.1 Cloning of amplified 16S rRNA genes

16S rRNA gene products amplified with the *Taq* polymerase were cloned using the TA cloning vector pCR4-TOPO (Invitrogen) and those amplified with the Phusion and Herculase polymerases were cloned using the pCR4Blunt-TOPO and the Zero Blunt cloning kit (Invitrogen), according to the manufacturer's instructions.

For both cloning kits 2 µl of PCR product was combined with 1 µl salt solution (Invitrogen), 2 µl dH_2O (Sigma-Aldrich) and 1 µl of the appropriate TOPO (TA or Blunt) cloning vector (Invitrogen, Paisley, UK). The solution was incubated for 30 minutes at room temperature, placed on ice and 2 µl of the cloning solution added to TOP10 chemically competent cells and after gentle mixing, incubated for 10 min on ice. To stop the transformation reaction cells were heat shocked at 42 °C for 30 sec on a heating block and immediately placed on ice. Aliquots of 250 µl SOC medium (Invitrogen) were added and the cells incubated for 1 h at 37 °C, while shaking at 250 rpm in an orbital shaker. Finally, the cell suspension was spread in 50 µl and 100 µl aliquots on pre-warmed Luria Bertani (LB) agar supplemented with 50 µg / ml kanamycin (Gibco/Invitrogen) and incubated aerobically at 37 °C over night.

2.3.11.2 M13 PCR of clone inserts

Clone colonies were touched with the end of a sterile 10 µl pipette tip and the cells suspended in 50 µl sterile water. One µl of the suspension was used as the template in a PCR reaction using Thermoprime *Taq* polymerase as described above with 5 pmol each of M13 Forward and M13 Reverse primers. Initial denaturation was run at 95 °C for 15 min. This was followed by 30 cycles of denaturation at 95 °C for 45 sec,

annealing at 55 °C for 45 sec and extension at 72 °C for 90 sec, with a final period of extension at 72 °C for 5 min.

2.3.12 Restriction length polymorphism

2.3.12.1A. naeslundii/S. sanguinis mix

Amplicons were digested with the restriction enzyme *Ase*I (10 µl PCR product, 7 µl sterile dH_2O, 2 µl NEbuffer 3 and 1 µl *Sty*I [NEB]) at 37 °C for 16 h. Two µl of the digest was subjected to electrophoresis on a 1 % agarose gel.

2.3.12.2B. dentium/S. sanguinis mix

Amplicons were digested with the restriction enzyme *Sty*I (10 µl PCR product, 6.8 µl sterile dH_2O, 2 µl NEbuffer 3, 0.2 µl BSA and 1 µl *Sty*I [NEB]) at 37 °C for 16 h. Two µl of the digest was subjected to electrophoresis on a 1 % agarose gel.

2.3.13 Comparison of Thermoprime, Herculase and Phusion polymerases for their ability to amplify low and high G+C bacterial templates

Broth cultures of *A. naeslundii* and *S. sanguinis* were adjusted to turbidities equivalent to 10^7 cfu / ml and mixed. DNA was extracted from the mixture by means of the phenol-chloroform and GenElute™ methods. In addition, equal amounts of DNA (18 µg /µl) from the two organisms were mixed. The 16S rRNA genes from the two preparations were amplified using primers 27F CM and 1492R using the Thermoprime *Taq*, Phusion and Herculase polymerases. Amplified genes were cloned using the appropriate cloning kit for each polymerase. Cloned genes were identified using a restriction enzyme assay as described above. All experiments were performed in triplicate.

2.3.14 Cell quantitation

Aliquots (600 µl) of overnight cultures of *A. naeslundii* and *S. sanguinis* were sonicated three times for 10 sec using a Vibra Cell™ sonicator (Sonics & Materials Inc., Newtown, CT, USA) at an amplitude of 40 units. Between each period of sonication, the cells were cooled on ice for 2 min. The cell suspensions were diluted in phosphate buffered saline (PBS, Dulbecco A, Oxoid, Basingstoke, UK) in 10-fold steps and 5 µl aliquots were examined microscopically in a Helber counting chamber with Thoma ruling at a magnification of 400× under phase contrast illumination. As a guide, dilutions resulting in 30 - 80 cells per big square of the Helber chamber were used for quantitation. It

proved impossible to distinguish the number of bacteria in clumps; if present, each clump was counted as if it were a single bacterium. Three slides were prepared for each culture, six large squares were counted separately going from I to VI as indicated in Figure 2.1 and the counts averaged. The number of microscopic counts of the bacterial suspensions were calculated using Equation 1 below and adjusted with PBS to get the colony forming unit count to $1.39 \times 10^9 \pm 2.3 \times 10^8$ in both cultures.

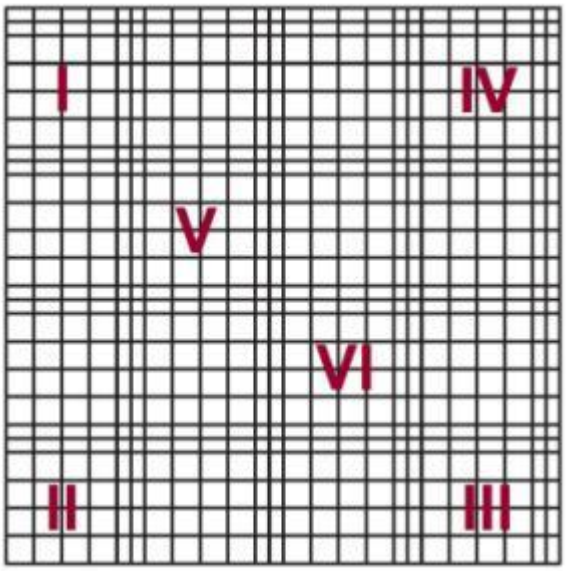


Figure 2.1: Schematic representation of Thoma ruling in a Helber counting chamber resulting in 16 big squares subdivided into 16 small squares.

$$CN/ml = \frac{\sum BS}{X} \times CF \times DF \quad \text{..... (1)}$$

Equation 2.1: CN/ml: cell number/ml, $\sum$ BS: number of cells per large square, X: number of big squares counted, DF: dilution factor, CF: chamber factor = $\frac{1000\mu l}{(depth \times area \times 16)}$

2.3.15 Comparison of Thermoprime and Phusion polymerases and three primer pairs for their ability to amplify low and high G+C bacterial templates

Broth cultures of *A. naeslundii* and *S. sanguinis* were adjusted to turbidities equivalent to $1.39 \times 10^9 \pm 2.3 \times 10^8$ cfu/ml and mixed in equal parts. DNA was extracted from the mixture by means of the GenEluteTM method and 16S rRNA genes were amplified

using primer pairs 27F / 1492R, 63F / 1387R and 39F / 1387R using the Thermoprime *Taq* and Phusion polymerases. Amplified genes were cloned using the appropriate cloning kit for each polymerase and cloned genes were identified as described above. All experiments were performed in triplicate.

2.4 Results

2.4.1 Identification of *A. naeslundii* and *S. sanguinis* by restriction fragment analysis of amplified 16S rRNA genes

Digestion of amplified cloned 16S rRNA genes from pure cultures of *A. naeslundii* and *S. sanguinis* with restriction enzyme *Sty*I gave distinctive banding patterns for each species (Figure 2.2). Due to the insertion orientation, two versions of banding patterns were observed. The inserted *A. naeslundii* DNA could either be cut into fragments of 390 bp, 558 bp and 714 bp by the restriction enzyme when inserted in one direction or in fragments of 374 bp, 573 bp and 714 bp when inserted in the other direction. The inserted *S. sanguinis* DNA could be cut into fragments of 379 bp and 1292 bp or 364 bp and 1307 bp.

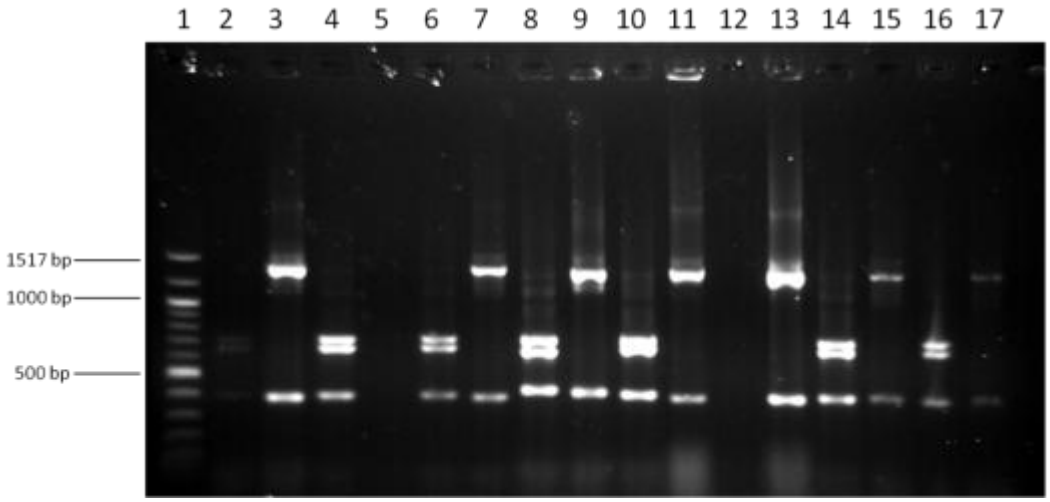


Figure 2.2: *Sty*I restriction fragment analysis of cloned 16S rRNA genes from *A. naeslundii* and *S. sanguinis*. Lane 1: DNA fragment length markers; lanes 2,4,6,8,10,12,14 and 16: *A. naeslundii* clones (H1 1-8); lanes 3,5,7,9,11,13,15 and 17: *S. sanguinis* clones (H2 1-8)

2.4.2 Comparison of different DNA extraction protocols for the amplification of 16S rRNA genes from Gram-positive bacteria with DNA of low and high G+C content

Five different DNA extraction protocols were tested for their ability to extract DNA from *S. sanguinis* and *A. naeslundii*. These comprised a phenol-chloroform protocol that has been widely used in the past, the UltraClean Microbial DNA isolation kit (Mo Bio), the FastDNA kit in conjunction with the FastPrep 24 homogenizer (MP BIO), the microLYSIS extraction solution (Microzone) and the GenEluteTM bacterial genomic

DNA kit (Sigma-Aldrich). Use of the MOBIO extraction kit was abandoned due to it introducing a PCR inhibitor.

The GenEluteTM kit, FastPrep kit and microLYSIS solution were tested on *A. naeslundii* and *S. sanguinis* broth cultures that were adjusted to specific cell densities using the calibration curves established with viable counts. Subsequent amplification using Thermoprime *Taq* and primer pair 27F CM / 1492R showed no extraction protocol resulted in equal signal strength for *A. naeslundii* and *S. sanguinis* in direct comparison (Figure 2.3). The microLYSIS protocol did result in far weaker signal strength for PCR products for *A. naeslundii* compared to the GenEluteTM and FastPrep kits and its use was discontinued.

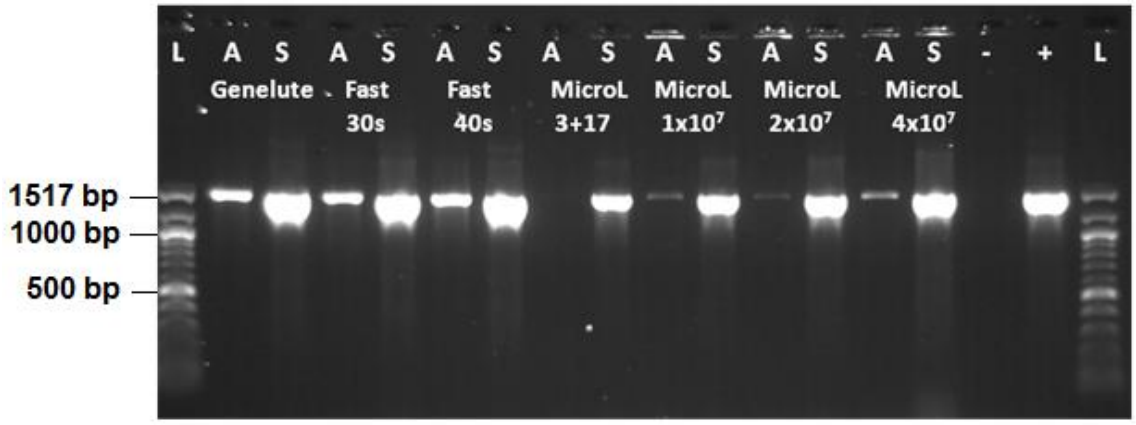


Figure 2.3: *Taq* 27F CM/1492R PCR amplification of DNA extracted from *A. naeslundii* (A) and *S. sanguinis* (S) using GenEluteTM, FastDNA and microLYSIS DNA extraction kits. FastDNA extraction using 30 or 40 *sec* homogenization. MicroLYSIS solution: 3µl cells ($2x10^8$ cfu/ml) + 17µl microLYSIS, 50µl ($1x10^7$) cells + 20µl microLYSIS, 100µl ($2x10^7$) cells + 20µl microLYSIS, 200µl ($4x10^7$) cells + 20µl microLYSIS, - water control, + *E. coli* positive control, L 100bp ladder

Further DNA extractions using the FastDNA kit and subsequent PCR amplifications gave inconsistent results with both Thermoprime and Phusion polymerases and use of this kit was discontinued.

2.4.3 Comparison of *Taq* and Pfu polymerases for the amplification of 16S rRNA genes from Gram-positive bacteria with DNA of low and high G+C content

The proportion of clones identified as *A. naeslundii* from libraries constructed from 16S rRNA genes amplified with the three polymerases from either DNA extracted from 50:50 mixtures of *A. naeslundii* and *S. sanguinis* cells by a phenol-chloroform

extraction method or the GenElute™ kit, or DNA from the two species mixed 50:50 (GenElute™ only), is shown in Table 2.3. When the cells of the two species were mixed in equal proportions, only two *A. naeslundii* clones were detected among the 540 clones identified. When the DNA was mixed in equal proportions prior to PCR, more *A. naeslundii* were seen in the libraries with 10 % (Phusion polymerase), 11.1 % (Herculase) and 17.7 % (Thermoprime) being identified as that species.

Table 2.3: Proportion of clones identified as *A. naeslundii* from libraries constructed from 16S rRNA genes amplified with Herculase, Phusion and Thermoprime polymerases from DNA extracted from 50:50 mixtures of *A. naeslundii* and *S. sanguinis* cells by phenol-chloroform method or the GenEluteTM commercial kit, or DNA from the two species mixed 50:50. Primers 27F CM and 1492R were used in all amplification reactions.

DNA extraction method	PCR template	Polymerase	No. of clones identified as *A. naeslundii* (n=30) Experiment			Mean proportion of *A. naeslundii* in library (%)
			1	2	3	
Phenol-chloroform	50:50 cell mix	Herculase	0	0	1	1.1
		Phusion	0	0	0	0
		Thermoprime	0	0	0	0
GenEluteTM	50:50 cell mix	Herculase	0	0	0	0
		Phusion	0	0	0	0
		Thermoprime	0	1	0	1.1
	DNA mix 50:50	Herculase	1	9	0	11.1
		Phusion	3	2	4	10.0
		Thermoprime	5	3	8	17.7

2.4.4 Cell quantitation

In order to determine if the disparity noted in the previous experiment between the results from mixtures of cells and mixtures of DNA was due to experimental error, the possibility that the initial quantitation of cell numbers had been inaccurate was investigated. No correlation between the absorbance readings and viable counts of five overnight cultures of *S. sanguinis* (correlation coefficient r=0.234) and overnight cultures of *A. naeslundii* (correlation coefficient r=0.0004) was found.

Microscopic examination of the broth cultures directly and Gram stains of the broth cultures revealed the presence of long chains and large clumps of *Actinomyces* that were not dispersed by vortex mixing by up to 90 sec. In the following experiments *S. sanguinis* was chosen because the long chains it forms are more difficult to disperse than the clumps formed by *A. naeslundii.* To test other methods of dispersion, 5 ml *S. sanguinis* broth culture was passed 5 times through a 25G needle and aliquots in appropriate dilutions were spread on BA to determine viable counts. However, when Gram-stained smears were examined post syringe treatment, it was observed that treated cultures still showed relatively long chains, demonstrating that cell numbers would still be underestimated. Finally, 600 µl *S. sanguinis* broth culture was sonicated at an amplitude of 40 for different time intervals. It was found that sonicating the broth culture for 3 x 10 sec interspersed with 2 min intervals on ice resulted in mostly individual cells with some diplococci; compared to mostly diplococci when sonicating for 10 sec once or twice and cell chains and debris when sonicating for 20 sec once or twice respectively. It was established subsequently that this protocol was also optimal for *A. naeslundii* broth cultures. The effect of sonicating the cultures three times for 10 sec on microscopic and viable counts is shown in Figure 2.4.

Microscopic quantitation using a Helber counting chamber when compared to viable counts spread on blood agar resulted in $6.6 \times 10^7 \pm 6.58 \times 10^6$ cfu / ml and $4.3 \times 10^7 \pm 1.77 \times 10^7$ cfu / ml, respectively, in the case of *S. sanguinis* cell units. For *A. naeslundii* these values were $6.02 \times 10^8 \pm 6.96 \times 10^7$ cfu / ml and $6.05 \times 10^8 \pm 2.57 \times 10^8$ cfu / ml, respectively (Figure 2.4).

The vortex mixed sample of *S. sanguinis* resulted in 52.1 % higher microscopic counts compared to viable counts. For the sonicated sample of *S. sanguinis* the same trend

was observed, the microscopic counts were 73.9 % higher than the viable counts. In the case of the vortex mixed sample of *A. naeslundii* microscopic counts were 4.9 % lower than the viable counts, but the standard deviation of the microscopic counts was 3.7 times lower than that of the viable counts. For the sonicated sample of *A. naeslundii* the same trend was observed as for the *S. sanguinis* samples, the microscopic counts were 57 % higher than the viable counts.

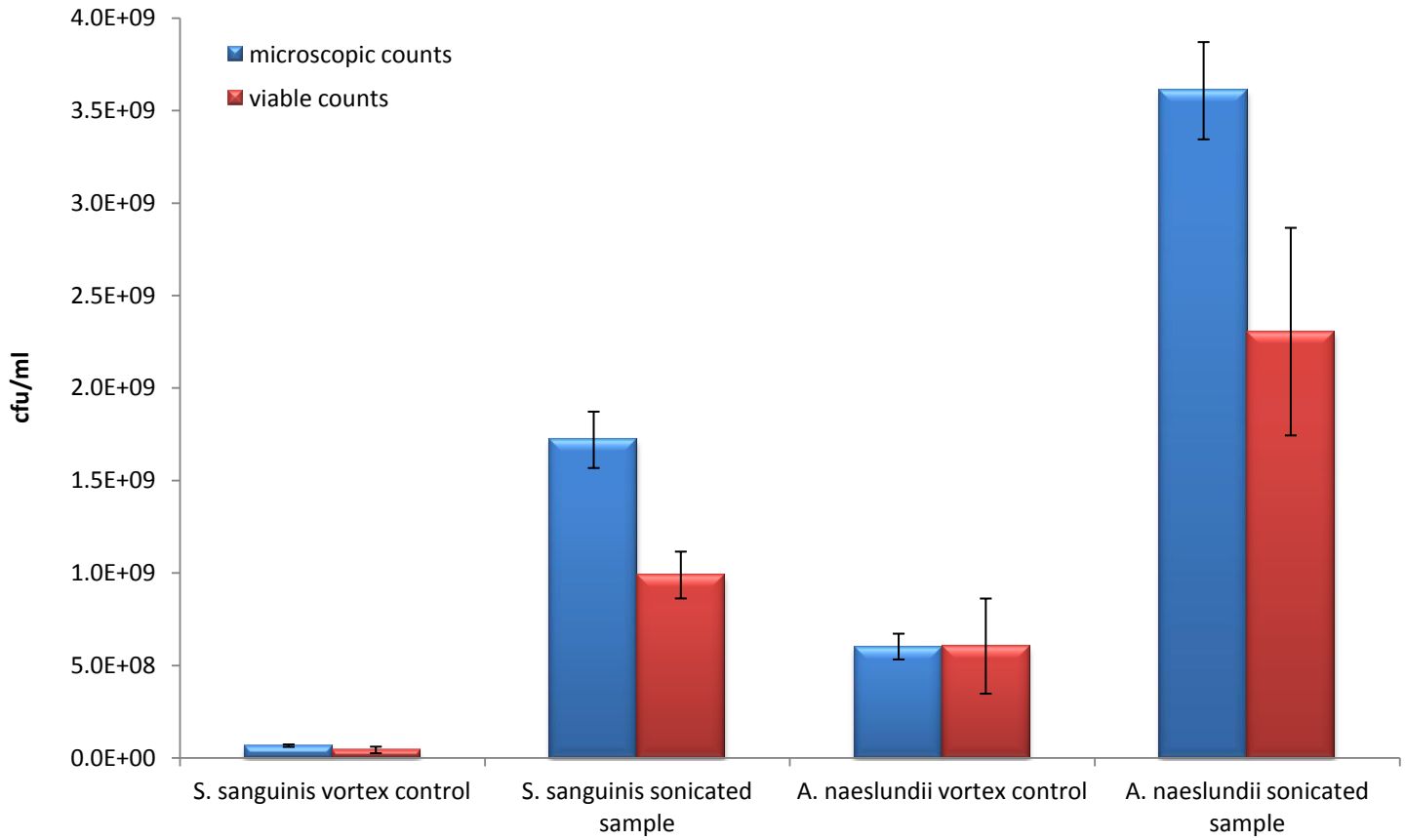


Figure 2.4: Microscopic and viable counts of *S. sanguinis* and *A. naeslundii* cultures treated by vortex mixing for 10 sec or 3× 10 sec sonication. Error bars showing standard deviation.

2.4.5 Comparison of "universal" primer sets 27F CM/1492R and 63F/1387R for their ability to amplify 16S rRNA genes from members of the phylum *Actinobacteria*

Primer set 27F CM / 1492R produced strong amplification from four of the five strains tested, but no PCR product was obtained from *B. dentium* (Table 2.4). In contrast, 63F / 1387R successfully amplified 16S rRNA genes in all strains, except *S. sanguinis*.

Table 2.4: Amount of PCR product obtained from amplification of 16S rRNA genes from oral bacterial species with primer sets 27F / 1492R and 63F /1387R using Thermoprime Taq polymerase. - no product; + weak bands; ++ moderate bands; +++ strong bands (determined by visual inspection according to brightness and size of the band).

Species	Forward primer	Reverse primer	Amount of PCR product
A. naeslundii	27F CM	1492R	+
A. parvulum			++
B. dentium			-
O. uli			+
S. sanguinis			+++
A. naeslundii	63F	1387R	+
A. parvulum			+++
B. dentium			++
O. uli			+++
S. sanguinis			-

The failure of 27F CM / 1492R to amplify the 16S rRNA gene in *B. dentium* was investigated further (Figure 2.5). Primer 27F CM produced only an extremely weak product when used with 1387R while 63F with both 1387R and 1492R gave good product.

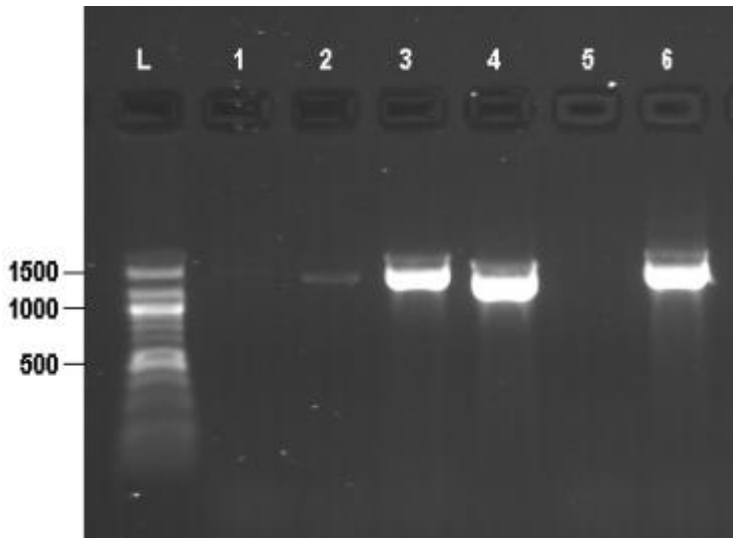


Figure 2.5: PCR amplification of 16S rRNA gene of *B. dentium*. Lane L - 100 bp ladder; lane 1 - 27F CM+1492R; lane 2 - 27F CM+1387R; lane 3 - 63F+1492R; lane 4 - 63F+1387R; lane 5 - negative water control; lane 6 - *E. coli* template, 27F+1492R.

Alignment of primer 27F CM with the 16S rRNA gene sequence of *B. dentium* revealed 3 mismatches (Figure 2.6).

```
27F           5′- AGAGTTTGATCMTGGCTCAG-3′
B. dentium    5′- ..G...C...TC........-3′
```

Figure 2.6: Nucleotide sequence alignment of primer and corresponding region in *B. dentium*.

The subsequent PCR using the newly designed species-specific forward primer bif27F CM, in which the three mismatched bases were replaced with those found in *B. dentium*, resulted in strong product bands in combination with reverse primer 1387R as well as 1492R (Table 2.5 and Figure 2.7).

Table 2.5: Amount of PCR product obtained from amplification of 16S rRNA genes from *B. dentium* with primer sets bif27F / 1387R and bif27F / 1492R and Thermoprime Taq polymerase- no product; + weak bands; ++ moderate bands; +++ strong bands.

Species	Forward primer	Reverse primer	Amount of PCR product
B. dentium	bif27F CM	1387R	+++
B. dentium	bif27F CM	1492R	+++

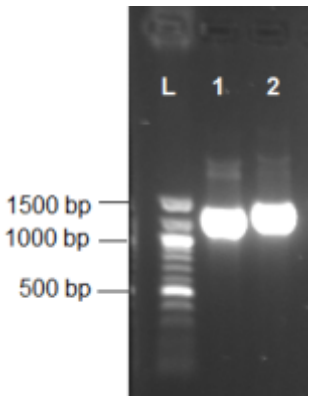


Figure 2.7: PCR amplification of 16S rRNA gene of *B. dentium*. Lane L - 100 bp ladder; lane 1 - bif27F CM+1387R; lane 2 - 27F CM+1492R

2.4.6 Evaluation of novel and modified 16S rRNA gene PCR primers

Two additional forward primers were designed and evaluated. 61F had the same sequence as 63F but with two bases removed from the 3´ end to eliminate the mis-match with members of the genus *Streptococcus*. Primer 39F is a new primer based on a conserved region of the 16S rRNA gene revealed by aligning representatives of all oral genera in the domain Bacteria (Weisburg et al. 1991; Baker et al. 2003). The results obtained using these primers with *Taq* polymerase are shown in Table 2.6. Primer pair 61F and 1387R gave strong or enhanced amount of product with all 5 oral species tested. Primer pair 61F / 1492R gave similar results but the product obtained from *A. naeslundii* was weak. Primer 39F with either 1387R or 1492R gave strong or better product with all strains. Excellent product was obtained for all primer combinations when *E. coli* was used as template.

Table 2.6: Amount of PCR product obtained from amplification of 16S rRNA genes from oral bacterial species with various primer sets and Thermoprime Taq polymerase. - no product; +/- very weak bands; + weak bands; ++ moderate bands; +++ strong bands.

Species	Forward primer	Reverse primer	Amount of PCR product
A. naeslundii	27F CM	1492R	++
A. parvulum			+++
B. dentium			+/-
O. uli			+

Species	Forward primer	Reverse primer	Amount of PCR product
S. sanguinis			+++
A. naeslundii	63F	1387R	++
A. parvulum			+++
B. dentium			+++
O. uli			+++
S. sanguinis			(+/-) -
A. naeslundii	61F	1387R	+
A. parvulum			++
B. dentium			++
O. uli			++
S. sanguinis			++
A. naeslundii	61F	1492R	+/-
A. parvulum			++
B. dentium			+
O. uli			++
S. sanguinis			+
A. naeslundii	39F	1387R	+++
A. parvulum			+++
B. dentium			+++
O. uli			+
S. sanguinis			+++

Species	Forward primer	Reverse primer	Amount of PCR product
A. naeslundii	39F	1492R	++
A. parvulum			+++
B. dentium			++
O. uli			+
S. sanguinis			+++
E. coli	27F CM	1492R	+++
E. coli	63F	1387R	+++
E. coli	61F	1387R	+++
E. coli	61F	1492R	+++
E. coli	39F	1387R	+++
E. coli	39F	1492R	+++

Table 2.7 shows the results of the same amplifications when Phusion polymerase was used. All primer combinations generated excellent PCR product.

Table 2.7: Amount of PCR product obtained from amplification of 16S rRNA genes from oral bacterial species with various primer sets and Phusion polymerase. . - no product; +/- very weak bands; + weak bands; ++ moderate bands; +++ strong bands.

Species	Forward primer	Reverse primer	Amount of PCR product
A. naeslundii	27F CM	1492R	+++
A. parvulum			+++
B. dentium			+++
O. uli			+++
S. sanguinis			+++

Species	Forward primer	Reverse primer	Amount of PCR product
A. naeslundii	63F	1387R	+++
A. parvulum			+++
B. dentium			+++
O. uli			+++
S. sanguinis			+++
A. naeslundii	39F	1387R	+++
A. parvulum			+++
B. dentium			+++
O. uli			+++
S. sanguinis			+++
A. naeslundii	39F	1492R	+++
A. parvulum			+++
B. dentium			+++
O. uli			+++
S. sanguinis			+++
A. naeslundii	61F	1387R	+++
A. parvulum			+++
B. dentium			+++
O. uli			+++
S. sanguinis			+++
A. naeslundii	61F	1492R	+++

Species	Forward primer	Reverse primer	Amount of PCR product
A. parvulum			+++
B. dentium			+++
O. uli			+++
S. sanguinis			+++
E. coli	27F CM	1492R	+++
E. coli	63F	1387R	+++
E. coli	39F	1387R	+++
E. coli	39F	1492R	+++
E. coli	61F	1387R	+++
E. coli	61F	1492R	+++

The amplification of 16S rRNA genes from oral bacterial species with various primer sets and Herculase polymerase gave inconsistent results in repeat experiments and the use of this polymerase was discontinued.

2.4.7 Comparison of three primer pairs and *Taq* and Pfu polymerases for the amplification of 16S rRNA genes from Gram-positive bacteria with DNA of low and high G+C content

In the following experiment primer pair 27F CM / 1492R was chosen as a reference point, since this primer is most commonly used in other studies targeting the 16S rRNA gene. Primer pair 39F / 1387R was chosen as it was designed to target oral species and resulted in strong PCR product for all tested species except for *O. uli.* The last primer pair, 61F / 1387R, was included as it promised improvement in the detection rate of *Streptococci.* The proportion of clones identified as *A. naeslundii*, from libraries constructed from 16S rRNA genes amplified with the two polymerases Thermoprime *Taq* and Phusion and three primer sets from DNA extracted from 50:50 mixtures of *A. naeslundii* and *S. sanguinis* cells by a commercial kit, is shown in Table 2.8. The primer

set 27F CM / 1492R that is used commonly in molecular microbial ecology studies resulted in the lowest value of detected *A. naeslundii* with 10.4 % for Thermoprime polymerase. The primer set 39F / 1387R detected 51 clones out of 135 (37.8 %) and the primer pair 61F / 1387R gave the highest result of 71.1 % with the same polymerase. The detection rate for *A. naeslundii* was lower when the same primer pairs were used with Phusion polymerase. The primer combinations 27F CM / 1492R, 39F / 1387R and 61F / 1387R identified 8.1 %, 16.3 % and 40.7 %, respectively, of the clones as *A. naeslundii*.

Table 2.8: Proportion of clones identified as *A. naeslundii* from libraries constructed from 16S rRNA genes amplified with Thermoprime and Phusion polymerases from DNA extracted from 50:50 mixtures of *A. naeslundii* and *S. sanguinis* cells by the GenEluteTM commercial kit. Primers 27F CM, 39F, 61F, 1387R and 1492R were used in the amplification reactions

Polymerase	Primer combination	No. of clones identified as *A. naeslundii* (n=45)			Mean proportion of *A. naeslundii* in library (%)
		Experiment			
		1	2	3	
Thermoprime	27F CM / 1492R	6	5	3	10.4
Thermoprime	39F / 1387R	14	21	16	37.8
Thermoprime	61F / 1387R	32	36	28	71.1
Phusion	27F CM / 1492R	5	4	2	8.1
Phusion	39F / 1387R	5	8	9	16.3
Phusion	61F / 1387R	20	19	16	40.7

2.4.8 Effect of reverse primer 1492R together with primer 27F CM and 39F on the amplification of 16S rRNA genes from Gram-positive bacteria with DNA of low and high G+C content using Thermoprime polymerase

The proportion of clones identified as *A. naeslundii*, from libraries constructed from 16S rRNA genes amplified with Thermoprime *Taq* and primer sets 27F CM / 1492R and 39F / 1492R from DNA extracted from 50:50 mixtures of *A. naeslundii* and *S. sanguinis* cells by the GenElute kit is shown in Table 2.9. In the previous experiment primer pair 39F / 1387R in conjunction with Thermoprime polymerase resulted in nearly 4x the detection rate of *A. naeslundii* compared to primer pair 27F CM / 1492R using the same polymerase. To be able to directly compare the influence of the forward primer the experiment was repeated using the same reverse primer for both forward primers. Primer set 27F CM / 1492R detected 4.3 % *A. naeslundii,* while Primer set 39F / 1492R resulted in a detection rate of 6.9 %.

Table 2.9: Proportion of clones identified as *A. naeslundii* from libraries constructed from 16S rRNA genes amplified with Thermoprime polymerase from DNA extracted from 50:50 mixtures of *A. naeslundii* and *S. sanguinis* cells by the GenEluteTM commercial kit. Primers 27F CM, 39F and 1492R were used in the amplification reactions

Polymerase	Primer combination	No. of clones identified as *A. naeslundii* (n=39) Experiment			Mean proportion of *A. naeslundii* in library (%)
		1	2	3	
Thermoprime	27F CM / 1492R	1	1	3	4.3
Thermoprime	39F / 1492R	4	1	3	6.9

2.4.9 Comparison of primer pairs 61F/1387R and 61F/1492R with Phusion polymerase for the amplification of 16S rRNA genes from Gram-positive bacteria with DNA of low and high G+C content

The proportion of clones identified as *A. naeslundii* from libraries constructed from 16S rRNA genes amplified Phusion polymerase and primer sets 61F / 1387R and 61F /

1492R from DNA extracted from 50:50 mixtures of *A. naeslundii* and *S. sanguinis* cells by the GenElute kit, is shown in Table 2.10. Primer set 61F CM / 1387R detected 36.8 % *A. naeslundii,* while a drop in detection of *A. naeslundii* was observed when using primer set 61F / 1492R, resulting in a detection rate of 16.2 %.

Table 2.10: Proportion of clones identified as *A. naeslundii* from libraries constructed from 16S rRNA genes amplified with Thermoprime polymerase from DNA extracted from 50:50 mixtures of *A. naeslundii* and *S. sanguinis* cells by the GenEluteTM commercial kit. Primers 61F, 1387R and 1492R were used in the amplification reactions

Polymerase	Primer combination	No. of clones identified as *A. naeslundii* (n=39)			Mean proportion of *A. naeslundii* in library (%)
		Experiment			
		1	2	3	
Phusion	61F / 1387R	13	16	14	36.8
Phusion	61F / 1492R	4	7	8	16.2

2.4.10 Comparison of primer pairs 39F / 1387R and 61F / 1387R and *Taq* and Pfu polymerases for the amplification of 16S rRNA genes from the Gram-positive bacteria *A.* naeslundii and *B.* dentium with DNA of low and high G+C content

In order to determine if the results from the above experiments using a mix of *A. naeslundii* as the high G+C organism and *S. sanguinis* as the low G+C organism were applicable to other high G+C organisms, the experiment was repeated using *B. dentium* (high G+C) and *S. sanguinis* in a 50:50 DNA mix. The primer/polymerase combinations that came closest to detecting 50 % *A. naeslundii* were chosen for this experiment.

The proportion of clones identified as the high G+C species (*A. naeslundii* or *B. dentium)* from libraries constructed from 16S rRNA genes amplified with the two polymerases Thermoprime *Taq* and Phusion and two primer sets from DNA extracted from 50:50 mixtures of high G+C species and *S. sanguinis* cells by the GeneElute kit, is shown in Table 2.11.

Table 2.11: Proportion of clones identified as the high G+C species (A - *A. naeslundii* or B - *B.dentium*) from libraries constructed from 16S rRNA genes amplified with Thermoprime and Phusion polymerases from DNA extracted from 50:50 mixtures of high G+C organism and *S. sanguinis* (S) cells by the GenElute™ commercial kit. Primers 39F, 61F and 1387R were used in the amplification reactions

Polymerase	Species mix	Primer combination	No. of clones identified as *A. naeslundii* (n=44) Experiment			Mean proportion of high G+C species in library (%)
			1	2	3	
Thermoprime	AS	39F / 1387R	17	21	21	44.70
Thermoprime	BS	39F / 1387R	14	11	10	26.52
Phusion	AS	61F / 1387R	16	11	11	28.79
Phusion	BS	61F / 1387R	34	19	35	66.67

The combination of *Taq* polymerase and primer set 39F / 1387R detected 59 clones out of 132 (44.7 %) for the *A. naeslundii/S. sanguinis* mix and 35 clones out of 132 (26.5 %) for the *B.dentium/S. sanguinis* mix. The combination of Phusion polymerase with primer pair 61F / 1387R resulted in a detection rate of 28.8 % (38 out of 132) for the *A. naeslundii/S. sanguinis* mix and 66.7 % (88 out of 132) for the *B. dentium/S. sanguinis* mix.

2.4.11 Comparison of primers 27F CM, 27F YM and bif27F paired with primers 1492R or 1387R using *Taq* polymerases for the ability to amplify 16S rRNA genes from *B. dentium*

Frank et al. (2008) described a primer mix of four parts 27F YM, one part 27F-Bor (Borelia), one part 27F-Chl. (Chlamydia) and one part 27F-Bif. The sequence of primer 27F-Bif was exactly the same sequence as bif27F in this study. A PCR was set up to compare the ability of 27F CM, 27F YM and bif27F to amplify the 16S rRNA gene of *B. dentium* using reverse primers 1387R and 1492R.

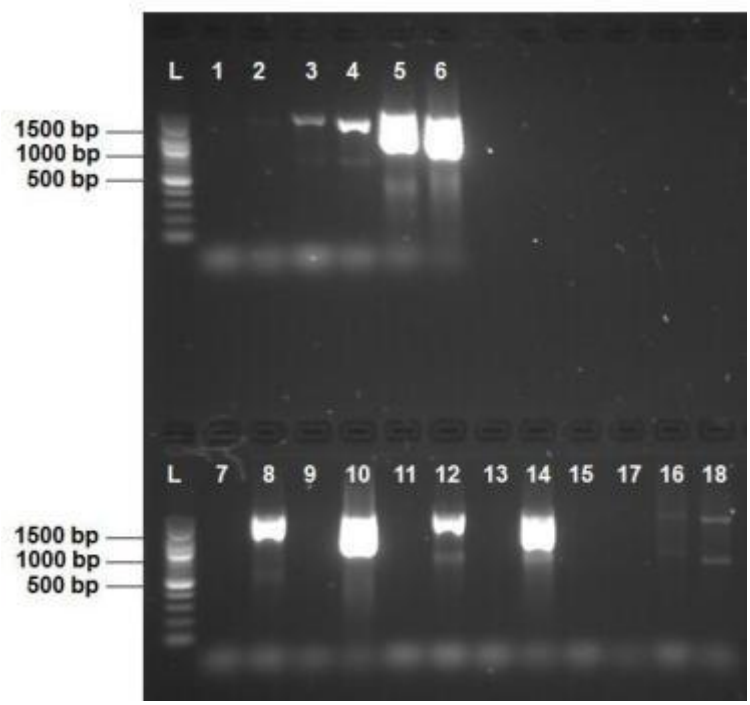


Figure 2.8: PCR amplification of 16S rRNA gene of *B. dentium*. Top row: lane L - 100 bp ladder; lane 1 - 27F CM+1492R; lane 2 - 27F CM+1387R; lane 3 27F YM+1492R; lane 4 27F YM+1387R; lane 5 bif27F+1492R; lane 6 bif27F+1387R; bottom row using water negative control and *E. coli* positive control: lane L - 100 bp ladder; lane 7 & 8 - 27F CM+1492R; lane 9 & 10 - 27F CM+1387R; lane 11 & 12 - 27F YM+1492R; lane 13 & 14 - 27F YM+1387R; lane 15 & 17 bif27F+1492R; lane 16 and 17 bif27F+1387R.

Primer combinations 27F CM / 1492R (lane 1) and 27F CM / 1387R (lane 2) did not result in any product. Reactions using 27F YM as the forward primer resulted in product, but the reaction using reverse primer 1492R (lane 3) gave less signal intensity than use of primer 1387R (lane 4). The strongest signal intensity was observed when forward primer bif27F was used together with either 1387R or 1492R (lane 5 and 6), as would be expected, since this primer is specific for this template. For the negative and positive control reactions water (lanes 7, 9, 11, 13, 15 and 17) and *E. coli* DNA (lanes 8, 10, 12, 14, 16 and 18) were used, respectively. The use of reverse primer 1387R (lanes 10, 14 and 18) here, too, resulted in stronger signal compared to primer 1492R (lanes 8, 12 and 16). Control reactions using *E. coli* DNA together with forward primer bif27F (lanes 15-18) and either reverse primer resulted in very weak bands, as would be expected, due to the specific nature of the forward primer for a different template, in this case *B. dentium*.

2.5 Discussion

It has been noted in a number of molecular ecology studies of the oral microbiota that members of the phylum *Actinobacteria* are under-represented in 16S rRNA gene clone libraries (Munson et al. 2002; Munson et al. 2004; de Lillo et al. 2006; Farris et al. 2007). Various factors could cause this. One of the first points of investigation in this chapter was the method of DNA extraction.

Use of the MoBio kit was discontinued as it resulted in the introduction of a PCR inhibitor, necessitating a DNA clean up step and therefore raising the potential of losing DNA and introducing further bias. The microLYSIS protocol was abandoned since PCR products showed weak signal strength for one of the two species tested. The FastDNA method was deemed unsuitable due to the low reproducibility in Thermoprime and Phusion polymerase PCR reactions. A study performing comparative analysis of microbial DNA extraction protocols for groundwater samples (Purswani et al. 2011) found the FastDNA kit DNA extractions consistent, but banding patterns on TGGE profiles were very different compared to the other three methods used and sequencing analysis revealed this kit to produce the lowest diversity. Of further concern using a bead beating approach, such as the FastDNA kit, is the potential of DNA shearing, rendering it redundant for downstream applications such as PCR and the extraction beads blocking the columns (Rantakokko-Jalava et al. 2002; de Boer et al. 2010; Purswani et al. 2011). As Rantakokko-Jalava (2002) observed, the ideal DNA extraction protocol should release DNA from all sample organisms equally efficiently, and wash out inhibitory factors without introducing contaminants. However, it should be borne in mind when analysing data that, to date, no single approach results in optimal recovery of DNA from the various bacterial species in clinical/environmental samples (Hendolin et al. 2000; Rantakokko-Jalava et al. 2002; Purswani et al. 2011). Extractions from pure cultures resulted in higher yields using the GenElute^{TM} kit (data not shown) as well as results comparable to phenol-chloroform when used for the comparison of *Taq* and Pfu polymerases for the amplification of 16S rRNA genes from Gram-positive bacteria with DNA of low and high G+C content. Phenol-chloroform furthermore necessitates use of dangerous chemicals such as phenol, which also carries the risk of inhibiting PCR reactions if not removed. Therefore the GenEluteTM bacterial genomic DNA kit was chosen as the optimal DNA extraction method.

The following set of experiments described here were designed to assess whether the polymerases and buffer systems designed for the amplification of high G+C DNA templates would improve the amplification of *Actinobacteria*. The results confirmed the bias against this group with *Taq* polymerase described previously (Varadaraj et al. 1994) but the Phusion and Herculase II polymerases provided no improvement in recovery, despite their processivity-enhancing domains that allow them to amplify high G+C DNA.

The accuracy of cell quantitation by adjusting the turbidity and performing viable counts was found to introduce a bias towards underestimating cell numbers. Validating microscopic cell counts performed using the Helber chamber method found that for both *A. naeslundii* and *S. sanguinis* cultures, vortex mixing did not separate cells from clumps and chains. Performing viable counts with these inadequately dispersed broth cultures resulted in chains and clumps of bacteria growing as one new colony rather than one individual cell forming one new colony. In summary, *S. sanguinis* counts post-sonication were on average 25 times higher than vortex control counts and *A. naeslundii* counts post sonication were on average five times higher than vortex control counts. The comparison of microscopic counts versus viable counts shows a similar trend in that *S. sanguinis* microscopic counts were on average 1.6 times higher than viable counts and *A. naeslundii* microscopic counts were 1.5 times higher than viable counts.

One disadvantage of the sonication process may be that it can kill viable cells (Olsen et al. 1981). This did not have an effect on the outcome of this experimental design however, since the Helber chamber counting method does not differentiate between live and dead cells and instead registers total cell counts. The rate of precision for microscopic count using the Helber chamber proved to be better than that of the viable counts in all cases, as can be deducted from the standard deviation shown as error bars in Figure 2.4. This again proved the Helber counting chamber method using sonication for dispersion to be superior to the viable count method.

The “universal” primer set 27F CM / 1492R produced no product in PCRs performed with *B. dentium* DNA as template. Subsequent alignment of the *B. dentium* 16S rRNA gene with the 27F CM sequence revealed three mismatches. Although these were at

the 5' end of the primer and might not have been expected to strongly influence amplification. However, use of a version of 27F CM matching the *B. dentium* sequence, bif27F, gave excellent product from the PCR. Interestingly, however, 27F CM has been used in previous studies where bifidobacteria have been detected, such as the characterisation of the microbiota associated with dentine caries described by Munson et al. (2004). Further evidence for the effect of primer design on detection of this group is provided by Becker et al. (2002) who used a reverse-capture checkerboard assay to determine the levels of 23 bacterial species in childhood caries, and found bifidobacteria to be among the predominant organisms present. In the initial PCR performed as part of the checkerboard procedure, Becker et al. (2002) used a modified version of 27F CM as shown below:

```
27F CM          5'-.AGAGTTTGATCMGGCTCAG-3'
B.dentium       5'-...G...C...TC.......-3'
27F (Becker)    5'-G..........YC.......-3'
```

Figure 2.9: Alignment of 27F CM with *B. dentium* sequence and primer 27F used by Becker et al.

This version has an additional guanine at the 5´end and, probably of greater importance, a wobble at position 12 that accommodates the mis-match between the original 27F CM sequence (C) and *B. dentium* (T).

The finding in this study, that primer 27F CM produced weak PCR product for many oral bacterial species confirmed the findings of the study by Marchesi et al. (1998), who also found that the combination of 63F / 1387R was more consistent in yielding a product than the primer pair 27F CM / 1392R. A possible explanation for this result, given by Marchesi et al. (1998), is that primer 27F CM may form an intramolecular duplex with a 5´ overhang that makes it susceptible to the 5'-3' exonuclease activity of the *Taq* polymerase.

PCR amplification performed with *S. sanguinis* using primer 63F in combination with 1387R generated no product. Noguchi et al. (2005), examining extraradicular biofilm-forming bacteria associated with refractory endodontic pathogens, did not detect any

members of the genus *Streptococcus* using the same primer pair. Marchesi *et al.* (1998) did, however, report successful amplification of the 16S rRNA gene from the type strain of *S. sanguinis*. The reason for this discrepancy is unclear but may relate to PCR conditions, taking into account variations among thermal cyclers and differences between different manufacturer's polymerases and buffer systems.

Since streptococci have a mismatch with the 3' terminal base in 63F, this primer was shortened by two bases at that end and designated 61F. This primer successfully amplified the 16S rRNA gene of *S. sanguinis* and amplification of all other species resulted in good product.

The final newly designed universal primer, 39F, in conjunction with either 1387R or 1492R gave very strong signals for all species, except *O. uli*, where the signal strength was weaker but still visible. These results were expected, since the primer 39F was designed from an alignment of representative oral bacterial species.

Primer 27F YM only has one mismatch with the *Bifidobacterium* sequence due to the two wobble positions rather than the single wobble position found in 27F CM and differs from bif27F only in the third position from the 5' end. It did amplify *B. dentium* 16S rRNA, although variation in signal strength was observed depending on choice of reverse primer. Frank et al. (2008) argued that mismatches with primer 27F CM probably cause a systematic underrepresentation of phylogenetic groups in rRNA gene libraries. To increase the yield of more representative results for the investigated habitat they used a cocktail of primers representing seven unique sequence variants. Since the *Bifidobacteriales* sequence requires a mispairing to bind to 27F YM, therefore potentially introducing bias, their primer mix also contained 27F-Bif. Results of this study confirm that primer 27F YM does not amplify *B. dentium* DNA as well as 27F-Bif / bif27F. Primer 27F YM does, however, result in relatively strong bands compared to use of primer 27F CM, especially when paired with reverse primer 1387R. Therefore, if the use of universal primers is desired, representation of *Bifidobacteriales* can potentially be improved by using primer 27F YM in conjunction with 1387R or 1492R.

The novel primers designed for this study have subsequently been extensively tested, following the recommendation by numerous sources that systematic empirical testing of newly designed primers, including optimisation of the PCR programme, is of

paramount importance (Marchesi et al. 1998; McPherson et al. 2000; Hughes et al. 2007). The amplification of various species with the *Taq* polymerase and different primer combinations show that the choice of primers greatly influences the outcome of any PCR reaction. A library constructed with 27F CM / 1492R, for example, would be likely to be depleted of *Bifidobacteriaceae*, whilst the primer combination 63F / 1387R discriminates against the genus *Streptococcus* (Noguchi et al. 2005; Frank et al. 2008).

However, the biases noted above appear to be polymerase-specific. The later experiments amplifying DNA from all five species with any of the other primer combinations using the Phusion polymerase resulted in the amplification of the 16S rDNA region and consistently produced very strong signals. Moreover, the Phusion polymerase could amplify template DNA with varying G+C contents reliably, which could indicate that this polymerase has a higher tolerance of mismatches than, for example, *Taq* polymerase. The failure to amplify certain DNA templates in this study, which were successfully amplified by others (Marchesi et al. 1998; Sakata et al. 2006) indicates that additional factors play an important role in the successful outcome of the reaction. It can be concluded from both unsuccessful and successful amplifications of template that factors, such as the chemical composition of PCR components, the PCR programme used and/or hardware elements (e.g. PCR cycler, PCR tubes etc.) may influence the polymerases and therefore play an important role in the amplification process.

The first experiment comparing three primer pairs and two polymerases (*Taq* and Pfu) for the amplification of 16S rRNA genes from Gram-positive bacteria with DNA from low and high G+C content showed that detection rates of *A. naeslundii* were superior for both primer sets 39F / 1387R and 61F / 1387R compared to 27F CM / 1492R independent of use of either polymerase. In the cases of Thermoprime + 39F / 1387R and Phusion + 61F / 1387R the detection rate of around 40 % came closest to the expected detection rate of 50 %.

To be able to compare directly the amplification using primer 39F to 27F CM / 1492R, the same reverse primer was used. In this case, the detection rate of *A. naeslundii* for 27F CM / 1492R was less than half of that in the previous set of experiments, and for 39F / 1492R the detection rate was 5.5 times lower than for 39F / 1387R. It had been

observed in some PCR amplifications that the use of primer 1387R resulted in stronger signal strength than use of primer 1492R when used with the same forward primer. There is detection variation between experiments and the impact of the reverse primer seems to have a compounding effect. Despite this, the trend that a higher level of *A. naeslundii* can be detected with 39F, namely 1.6 times compared to 27F CM, is evident.

Similar results were observed with the Phusion polymerase and 61F / 1387R and 61F / 1492R. Overall detection of *A. naeslundii* was slightly lower at 36.8 % compared to the initial experiment (40.7 %), but the use of reverse primer 1387R resulted in a 2.2 times higher detection rate of *A. naeslundii* compared to reverse primer 1492R.

This suggests that not just use of primer 27F CM can have a damaging effect on amplification rates of species with a high G+C content, but the same applies to reverse primer 1492R. This is supported by findings of a study looking at the detection of Actinobacteria cultivated from environmental samples, which revealed a bias in universal primers (Farris et al. 2007). More specifically, in this study two versions of 1492R failed to amplify DNA from Actinobacteria, while the same DNA resulted in PCR product when used with Actinobacteria-specific primers. Analysis of the samples successfully amplified with Actinobacteria-specific primers and the 1492R primer sequence revealed a perfect match between primer and target sequence with no obvious conformational problems (Farris et al. 2007). They speculated that genomic DNA outside the target region could inhibit amplification, as had been reported by Hansen et al. (1998). Repeating the experiment with the *A. naeslundii*/*S. sanguinis* mix using Thermoprime polymerase + primer pair 39F / 1387R and Phusion polymerase + primer pair 61F / 1387R, and also introducing another high G+C / low G+C mix in the form of *B. dentium / S. sanguinis*, confirmed the inter-experimental variations. The detection rate of *A. naeslundii* was 18.3 % higher for Thermoprime + 39F / 1387R compared to the initial experiment. For Phusion polymerase and primer pair 61F / 1387R, on the other hand, the detection rate dropped by nearly 30 %.

Furthermore, variation in the detection of the two high G+C species was observed. Whereas the *A. naeslundii* / *S. sanguinis* mix was again close to the desired 50 % value, the detection rate for *B. dentium* was much lower at 26.5 %. The reverse was observed

in the case of Phusion polymerase and primer pair 61F / 1387R. Not only was the detection of *A. naeslundii* lower in this experiment at 28.8 %, it also seemed that this combination introduces a bias towards the detection of *B. dentium* (66.7 %), much like that observed when the Thermoprime polymerase was used with 61F / 1387R for the detection of *A. naeslundii* (71.1 %)

2.6 Conclusions

The work described in this report confirms the findings of earlier studies in showing that the amplification of mixed species DNA of varying degrees of G+C content is biased against organisms with DNA of high G+C content. However, the data did not show that *Taq* polymerase was responsible for this bias but, rather, that primer design was responsible. It was confirmed that so-called "universal" PCR primers do not amplify all members of the target groups and that the choice of primers may greatly influence the outcome of the study. Careful choice and / or use of combinations of primers is therefore of paramount importance in molecular ecology studies.

Chapter 3

Molecular characterisation of the bacterial community in dentine caries using culture and Sanger based sequencing

3 Chapter 3

3.1 Introduction

Dental caries, or tooth decay, is the dissolution of tooth structure by acids formed by bacteria as a result of the fermentation of dietary carbohydrate, particularly sucrose. *Streptococcus mutans* was historically one of the first species associated with dental decay leading to a caries lesion in the tooth (Clarke 1924; Loesche et al. 1975). However, *S. mutans*-free caries lesions have been observed (Marsh et al. 1989) and it has been recognized that not individual species, but most likely mechanisms, such as lowering of the pH in the oral cavity and subsequent demineralisation of the tooth, are crucial in the initiation and progression of the caries lesion (Marsh 2003). Consequently, it is vital to establish knowledge of all species colonising the oral cavity enabling investigation as to which mechanisms and interdependencies of species, such as co-localisation, co-aggregation and/or symbiotic relationships, are responsible for the initiation and progression of dental decay. Only by considering these factors is it possible to develop preventative and/or therapeutic measurements for this disease.

Traditionally, microbiologists have used culture media to grow and characterise bacterial species, but it was recognised from the great plate count anomaly that not all species can be readily grown under laboratory conditions. Consequently, in recent years molecular methods targeting the 16S rRNA gene to characterise complex microbial communities have been established and many sequences representing novel species have been detected. In fact, approximately 280 species from the oral cavity have been cultivated and formally named, while it is estimated that there are between 500 and 700 oral species (Paster et al. 2001; Paster et al. 2006; Dewhirst et al. 2010)

However, various drawbacks of using molecular methods transpired, for example it was found the proportions of *Actinobacteria* were underestimated using molecular analysis when a direct comparison to culture was available (Munson et al. 2002; Munson et al. 2004; de Lillo et al. 2006). Furthermore, recent studies by Tanner et al. (2011) and Kanasi et al. (2010) observed greater diversity of species detected in early childhood caries (ECC) using culture compared to clonal analysis.

Even though many species have been detected in caries samples, most, if not all studies in the last few years have reported detection of potentially novel species,

genera or even higher taxonomic orders (Munson et al. 2004; Nadkarni et al. 2004; Chhour et al. 2005; Kanasi et al. 2010; Tanner et al. 2011). Tanner et al. (2011) reported *Scardovia wiggsiae* to be significantly associated with severe ECC children in the presence and absence of *S. mutans* detection and showed for the first time a strong association of *S. wiggsiae* together with *S. mutans* in ECC. These findings clearly demonstrate that continued efforts to characterize the microbiota of dentine caries and distinguish mechanisms of disease progression are needed.

In chapter 2 the problem of primer bias was addressed, focusing specifically on trying to enhance detection of high G+C species. A novel primer was designed, while another was adapted and those together with existing primers were validated and evaluated in various combinations on pure species and mock communities.

3.2 Aims

The aim of this chapter was to revisit the microbiology of dentine caries lesions and to further the knowledge of bacterial taxa, especially those with a high content of G+C, using a combination of culture and molecular (Sanger 16S rRNA gene sequencing) methods.

3.3 Materials and methods

3.3.1 Patient sample collection

Ethical approval for the study was granted by the Lewisham Local Research Ethics Committee South London REC Office (4) (Reference 08/H0810/61). Six patients, four male and two female, aged 22 to 35 years (mean age 26.6 years), who were medically fit and well participated in the study with their informed consent. Patients were included if they had a carious lesion that had spread into the middle or inner third of dentine, that was checked radiographically with cavitation. Local anaesthesia was administered where necessary, and the carious teeth isolated with rubber dam to minimize saliva contamination during the excavation procedure. Following removal of carious enamel to the enamel-dentine junction with a sterile, water-cooled diamond bur in an air-turbine handpiece, the dentine lesion was hand excavated with a sterile, spoon excavator (Ash G5; Claudius Ash Ltd., Potters Bar, UK). After the superficial layer of debris had been removed and discarded, the sample, consisting of soft necrotic dentine, was collected using a fresh, sterile spoon excavator at a level that represented the infected dentine lesion. This dentine was clinically soft to probe and discoloured (with varying levels of brown staining). The cavities were then lined if necessary and restored with a suitable restorative material.

3.3.2 Sample processing

Reduced transport medium (RTM) was prepared with the following composition: 1 % w/v tryptone, 0.5 % w/v yeast extract, 0.1 % w/v L-cysteine, 0.1 % w/v D+glucose, 2 % v/v horse serum to a total volume of 100 ml water and adjusted to pH 7.5. The medium was filter-sterilised (0.2 µm) and pre-reduced in the anaerobic workstation for 24 h. Each of the six carious dentine samples was deposited in a sterile 2 ml screw cap vial containing 1 ml of RTM and taken immediately to the laboratory. Samples were placed inside an anaerobic workstation and the vial cap opened for 30 sec to replace the gaseous headspace with the anaerobic atmosphere. Samples were then vortex-mixed for 1 min.

3.3.3 Culturing of sample organisms

Samples were diluted and plated inside the anaerobic workstation. Ten-fold serial dilutions were prepared in RTM. 100 µl of dilutions (10^{-3} to 10^{-7}) were used to

inoculate pre-reduced fastidious anaerobic agar (FAA; LabM) / 5% horse blood plates, in triplicate, and incubated anaerobically for 10 d in the anaerobic workstation at 37 °C. Colonies were selected for subculture at random by the following method. Plates with between 30 and 300 colonies were counted and the total number of colonies divided by the number of isolates required, namely 96. The bottom of the plate was divided into zones so as to give approximately 96 of colonies in each zone. One zone was then chosen at random, after which the plate was turned over and all the colonies in that zone were subcultured, moving methodically from the left to right hand side of the zone to the right, until the required number of colonies were selected. Subcultured isolates were plated out in a regressive streak on FAA plates, with a *Propionibacterium acnes* feeder streak. Isolates were incubated anaerobically for a further 4 - 5 days, after which the purity of all isolates was visually checked using a plate microscope. Mixed cultures were subcultured to achieve purity and the resulting pure cultures were stored at -70 °C in BHI + 10 % glycerol.

3.3.4 DNA extraction

3.3.4.1 DNA extraction from samples

The remaining 900 µl of the sample was centrifuged for 10 min at 13 000 g, the supernatant discarded and the pellet subsequently used for DNA extraction using the GenElute™ bacterial genomic kit (Sigma Aldrich) as described in section 2.3.5.

3.3.4.2 DNA extracted from isolates

Cells were harvested from FAA plates of the isolates, incubated sufficiently long to give good growth, and suspended in 1 ml phosphate buffered saline (PBS, Oxoid). DNA was extracted as described (section 2.3.5).

3.3.5 PCR amplification of 16S rRNA genes

3.3.5.1 Thermoprime Taq PCR

16S rRNA genes of the extracted DNA from each patient sample were amplified with five sets of primers: 27F YM / 1492R, 27F CM / 1492R, 39F / 1387R, 39F / 1492R and 61F / 1387R, as described in Chapter 2. Five replicate amplification reactions were set up for each sample and primer set as described in section 2.3.10.1. Initial denaturation was at 95 °C for 5 min, followed by 25 cycles of denaturation at 95 °C for 45 sec, annealing at 50 °C for 45 sec and extension at 72 °C for 90 sec.

3.3.5.2 Phusion High fidelity Hot Start PCR

The Phusion polymerase was used to amplify 16S rRNA genes from isolates with primer pair 27F CM / 1492R. Reactions were prepared as described (section 2.3.10.3). Initial denaturation was at 98 °C for 30 sec, followed by 30 cycles of denaturation at 98 °C for 10 sec, annealing at 56 °C for 30 s and extension at 72 °C for 45 sec.

3.3.5.3 Phusion High fidelity Hot Start PCR – touch PCR

Some isolates (*Anaeroglobus geminatus* OT 121, *Peptococcus* OT 167, *Peptostreptococcaceae* OT 091 and *Veillonellaceae* OT 155) were extremely slow growing and insufficient growth was obtained to perform the GenEluteTM DNA extraction. For these isolates, direct touch PCR was performed; briefly, colonies were touched with the end of a sterile 10 µl pipette tip and the cells suspended in 50 µl sterile water. One µl of the suspension was used as the template in a PCR reaction using Phusion polymerase as described (section 2.3.10.3). Cells were subjected to an initial boiling step at 95 °C for 10 min. Initial denaturation was at 98 °C for 30 sec, followed by 30 cycles of denaturation at 98 °C for 10 sec, annealing at 56 °C for 30 sec and extension at 72 °C for 45 sec. A final extension was run at 72 °C for 5 min before the reaction was held at 4 °C.

3.3.6 Alternative amplification of isolates

For some isolates, the amplicon yield from PCR was weak. For these isolates, DNA was instead amplified using primer pair 61F / 1387R with Phusion or Taq polymerases as indicated in Table 3.1, since this was the only primer pair resulting in PCR products for all samples that did not amplify using 27F CM / 1492R.

3.3.7 Cloning of *Taq* amplified 16S rRNA genes

Ten µl of the five replicate *Taq* polymerase PCR products of each primer set were pooled and cloned into the TA cloning vector pCR4-TOPO (Invitrogen) as described (section 2.3.11.1). Transconjugants were detected on LB agar supplemented with 50 µg / ml kanamycin.

3.3.8 Amplification of library clone inserts

Ninety-six clone colonies were chosen at random and the insert amplified by PCR with vector-specific primers M13 FWD and REV as described in section 2.3.11.2.

Table 3.1: Isolates amplified using an alternative protocol

Phusion 61F / 1387R				Taq 61F / 1387R	
B002	D012	F004	F049	B009	F044
B012	D020	F005	F069	B026	F070
B024	D021	F008	F071	B047	F074
B038	D040	F010	F075	B063	
C001	D047	F012	F076	B067	
C002	D048	F017	F081	B069	
C005	D051	F019	F086	B070	
C006	D077	F022	F087	D039	
C020	D082	F023	F089	D068	
C025	D096	F027	F090	D078	
C026	E077	F028	F093	D083	
C031	E081	F029		D085	
C092	F001	F033		C086	
C096	F002	F046		F026	

3.3.9 Storage of library clones

Aliquots (150 µl) of liquid Luria Bertani (LB) medium containing 10 % glycerol and 50 µg / ml Kanamycin was dispensed into the wells of a flat-bottomed 96 microtitre plate. The broth in each well was inoculated with individual clones by touching clones with a pipette tips and swirling the pipette tip in a well with LB medium. The plate was incubated aerobically over night at 37 °C and then stored at -70 °C.

3.3.10 PCR product purification

PCR products from clone insert amplification or from isolates amplified with Phusion polymerase were purified using the ExoSAP-IT (Exonuclease I/Shrimp Alkaline Phosphatase) clean up kit (Affymetrix, High Wycombe, UK). Five µl of PCR product were mixed with 1 µl ExoSAP-IT and 1 µl dH_2O and incubated in a thermal cycler for 15 min at 37 °C followed by heat inactivation of the enzymes at 80 °C for 15 min.

3.3.11 Sequencing of clones and isolates

For each sample, 96 isolates and 96 clones from each library were partially sequenced using the universal primer 519R (5' GWATTACCGCGGCKG 3'). Reactions were set up with 0.5 µl BigDye (Applied Biosystems, Life Technologies), 1.75 µl 5 x sequencing buffer (Applied Biosystems), 0.3 µl primer 519R (10 µM), 5.45 µl deionised, autoclaved water and 2 µl cleaned up PCR product as template. Thirty cycles were run consisting of 10 sec at 96 °C, 5 sec at 50 °C and 2 min at 60 °C.

Subsequent to amplification products were cleaned up by adding to each well 10 µl deionised autoclaved water, followed by 50 µl precipitation mix, prepared for a whole 96-well plate (7200 µl 99.6 % ethanol, 300 µl dH_2O, 300 µl 0.1 mM EDTA and 150 µl 3M sodium acetate). The sealed plate was vortex mixed for 15 sec and incubated at room temperature for 20 min. Following centrifugation at 2.879 g and 4 °C for 25 min the supernatant was removed by inverting the plates on paper tissue. The plates were then centrifuged inverted on blotting paper for 10 sec at 0.216 g and 100 µl ice-cold 70 % ethanol added. The plates were again spun at 2.879 g and 4 °C for 10 min after which the supernatant was again removed by inverting the plates on paper tissue. The plates were again centrifuged inverted on blotting paper for 10 sec at 0.216 g and dried under vacuum for 5 min. Finally, 10 µl 0.1x TE was added to each well, the plate sealed, carefully vortex mixed and pulsed down in the plate centrifuge. The sequencing analyser used was a AB3730xl DNA analyzer (Applied Biosystems).

3.3.12 *In silico* analysis of DNA sequencing results

Sequences were provisionally identified by a *BLAST (Basic Local Alignment Search Tool)* search on the Human Oral Microbiome Database (HOMD, www.homd.org). Sequences showing homology of less than 98.5 % to database reference sequences were subjected to full length sequencing of the 16S rRNA gene.

3.3.13 Full length sequencing

Sequencing was performed as described (section 3.3.11) using the sequencing primers shown in

Table 3.2. Primers A and J were exchanged with primers A*, A** or J* where appropriate.

Table 3.2: Primer sequences used for full length sequencing of isolates and clones

Label	Primer	Primer sequence [2]
A	27F CM	5´-AGAGTTTGATCMTGGCTCAG-3´
A*	39F	5´-ATCMTGGCTCAGRWYGAACGC-3´
A**	61F	5´-CAGGCCTAACACATGCAAG-3´
B	342R	5´-CTGCTGCSYCCCGTAG-3´
C	357F	5´-CTCCTACGGGAGGCAGCAG-3´
D	519R	5´-GWATTACCGCGGCKGCTG-3´
F	907R	5´-CCGTCAATTCMTTTRAGTTT-3´
G	926F	5´-AAACTYAAAKGAATTGACGG-3´
H	1100R	5´-GGGTTGCGCTCGTTG-3´
I	1114F	5´-GCAACGAGCGCAACCC-3´
J	1392R	5´-ACGGGCGGTGTGTRC-3´
J*	1387R	5´-GGGCGGWGTGTACAAGGC-3´
K	1492R	5´-TACGGYTACCTTGTTACGACTT-3´

3.3.14 Data analysis

The composition of the libraries and isolates was analysed initially using descriptive statistics. Good's coverage statistic (Good 1953) was calculated, where Good's Coverage = [1 - (*n*/N)] × 100 - where *n* is the number of molecular species represented by one clone (single-clone OTUs) and N is the total number of sequences. Heatmaps of the top 50 taxa for clones and isolates were created. Species richness and diversity estimates were made using the EstimateS suite of programmes (Colwell 2009). Chao1, an estimate of species richness, Inverse Simpson diversity index and ICE, an incidence-based coverage estimator were calculated. Library 1 and the combined data set of all sequences were computed using the recommended bias corrected option, while the isolates, libraries 2, 3, 4 and 5 were computed using the classic option on recommendation by Anne Chao, since the Chao's estimated for confidence value for abundance distribution was > 0.5 in these instances.

[2] IUPAC of degenerate DNA bases: M = C or A, Y = C or T, K = G or T, R = A or G, S = G or C, W = A or T

3.3.14.1 Phylogenetic trees

The full-length 16S rRNA gene sequences representing novel taxa were aligned with appropriate sequences from the HOMD database as well as oral and non-oral sequences from the Ribosomal Database Project. Sequences were aligned using BioEdit (Hall 2011) and positions with missing bases removed. Trees were constructed using MEGA version 5 (Tamura et al. 2011). Distance matrices were constructed by the method of Jukes and Cantor and trees derived using the neighbour-joining method of Saitou and Nei. Bootstrapping was performed using 1,000 resamplings.

3.3.14.2 Heatmaps

Heatmaps were compiled by sorting the genera of all libraries into highest to lowest detection levels. That table was then bisected into culture and molecular analysis and each table again sorted by highest to lowest ranking value.

3.3.14.3 Data analysis using mothur

The sequence data were additionally analysed using mothur, an open-source, platform-independent, community-supported software suite for describing and comparing microbial communities (Schloss et al. 2009). The following pipeline was used to analyse the data:

The data were de-replicated using unique.seqs, thus removing redundant sequences and thereby reducing processing time. Using align.seqs the sequences were then aligned to the silva.bacteria 16S rRNA reference file (Pruesse et al. 2007). Sequences that had more than five ambiguous bases, that did not start by position 1137 (97.5 %-tile) or end by position 9800 (2.5 % -tile) were removed using the screen.seqs command. The filter.seqs program was then used to remove any columns with a '-' in every sequence after which unique.seqs was run again to remove any redundant sequences. Pre.cluster was used to merge sequences that were within 1 bp per 100 bp of total sequence length of a more abundant sequence with that sequence. The dist.seqs program calculated uncorrected pairwise distances between aligned DNA sequences and the cluster command was used to assign sequences to OTUs. Following this, a table was created indicating the number of times an OTU was present in each sample using the make.shared command. The classify.otu command was used to get a consensus taxonomy for an OTU at a value of 98.5 % (or 0.015) using the HOMD

version 10 reference sequence and taxonomy databases. A phylip-formatted distance matrix was calculated using dist.seqs followed by the clearcut and libshuff commands. The collect.single command was used to calculate the Chao1 richness and the inverse Simpson diversity index, while the rarefaction.single command was used to compile rarefaction curve data. A table containing the number of sequences, sample coverage, number of observed OTUs and the invsimpson diversity estimate was compiled using the summary.single command.

3.3.14.4Statistical analysis

A two tailed Z-Test Calculator for paired comparisons was used with a significance threshold of 0.05.

3.4 Results

The count of colony forming units on the FAA plate for the culture analysis of sample A, from which 96 isolates were subcultured to achieve pure cultures, was 5.36 x 10^4 cfu/ml. The number of cfu/ml for patient B was 1.26 x 10^7 cfu/ml and 2.04 x 10^8 cfu/ml for patient C. The colony count for patient D resulted in 3.36 x 10^7 cfu/ml, while the sample of patients E and F produced counts of 2.64 x 10^7 cfu/ml and 1.15 x 10^8 cfu/ml, respectively.

3.4.1 Analysis

The data set was screened for the occurrence of chimeras using the Mallard program (Ashelford et al. 2006), which reported no detectable chimeras, however, 2 chimeras were spotted by visual inspection (0.062%).

Analysis of the 2700 clones and 540 isolates resulted in identification of 229 taxa at species level, representing 8 phyla: *Firmicutes*, *Proteobacteria*, *Actinobacteria*, *Bacteroidetes*, *Fusobacteria*, TM7, *Spirochetes* and *Synergistetes*. The number of taxa detected for each phylum and the numbers of novel taxa (i.e. one or more sequences with sequence identity less than 98.5 % to human oral taxa in the HOMD version 10 or extended set version 1 databases) and their corresponding libraries are shown in Table 3.3.

Of the 216 taxa detected using molecular analysis, 143 could only be found using this method and all of the 16 novel taxa were amongst these. Using culture methods 86 taxa were detected in total, none of which were novel and 12 of these were not detected using molecular analysis. As described in Table 3.3, 16 novel sequences were detected; ranging from species level to order level. All sequences provisionally identified as novel were sequenced in near full length and sequences deposited in GenBank of the National Center for Biotechnology Information , US National Library of Medicine, using the BankIT system. Sequences were submitted and corresponding accession numbers received as detailed in Table 3.4.

Table 3.3: Detected phyla, number of taxa within the phylum and number of novel taxa for each phylum. Library 1 – 27F CM / 1492R, library 2 – 27F YM / 1492R, library 3 – 39F / 1387R, library 4 – 39F / 1492R and library 5 – 61F / 1387R.

Phylum	No. of taxa identified	No. of novel taxa	Libraries in which novel taxa detected
Firmicutes	99	6	1, 2 & 5
Proteobacteria	21	2	1 & 5
Actinobacteria	22	1	3
Bacteroidetes	60	7	1, 2, 3, 4 & 5
Fusobacteria	10	0	-
TM7	5	0	-
Spirochetes	10	0	-
Synergistetes	1	0	-
Total	**228**	**16**	

The distribution of phyla found in the dentine caries lesion detected through culture and molecular analysis is shown in Figure 3.1 and Table 1 (Appendix 1). *Streptococcus mutans* and lactobacilli, the organisms traditionally associated with caries were seen in most samples but, overall, a highly diverse community was seen, including numerous representatives of the genus *Prevotella*, *Lactobacillus, Selenomonas* and *Streptococcus.* The five most abundant species were *Lactobacillus gasseri*, *P. denticola*, *P. tannerae*, *S. mutans* and *Streptococcus sp.* OT 070, representing nearly one third (31.6 %) of all detected and identified sequences. There were differences in the composition of the microbiota in samples from different subjects; e. g. samples patients A and E had a significantly higher proportion of lactobacilli ($p < 0.05$) compared to all other patients, with few other *Firmicutes* observed in patient A, while the sample from patient E showed a significantly higher proportion of *S. mutans* and *Veillonella dispar/parvula* ($p < 0.05$).

Table 3.4: Novel sequences and their corresponding accession numbers

Accession number	Name	Clone ID	Species	Genus
JQ406529	*Haemophilus* K1	AL159	√	
JQ406530	*Comamonadaceae* K1	BL117		√
JQ406531	*Erysipelotrichaceae* K1	BL163		√
JQ406532	*Prevotella* K1	BL176	√	
JQ406533	*Flavobacteriaceae* K1	BL190		√
JQ406534	*Prevotella* K2	BL216	√	
JQ406535	*Capnocytophaga* K1	BL218	√	
JQ406536	*Bacteroidales* K1	BL234		√[3]
JQ406537	*Veillonellaceae* K1	DL111	√	
JQ406538	*Erysipelotrichaceae* K2	DL151		√
JQ406539	*Prevotella* K3	DL303	√	
JQ406540	*Porphyromonodaceae* K1	DL325		√
JQ406541	*Veillonella* K1	EL130	√	
JQ406542	*Capnocytophaga* K2	FL163	√	
JQ406543	*Actinomyces* K1	FL377	√	
JQ406544	*Moryella* K1	CL133	√	

Patients B, C, D and F had virtually no lactobacilli, but a wide range of other *Firmicutes*. Patient C had significantly higher levels of *Atopobium rimae* and *Atopobium* OT416 compared to all other patients ($p < 0.05$) and 88 of the detected 92 *Pseudoramibacter alactolyticus* clones were detected in this patient sample ($p < 0.05$). In patients B and D, five taxa belonging to the genus *Porphyromonas* were detected, while in the other patients a maximum of two taxa belonging to this genus were detected (A=1, C=1, E=0 and F=2). Compared to the levels of lactobacilli seen, significantly fewer *Prevotella* species were detected in patient samples A and E ($p < 0.05$). Between 21.11 and 40.19 % of sequences of patient B (25.19 %), C (38.70 %), D (40.19 %) and F (21.11 %) were made up of *Prevotella*, which was significantly more compared to patients A and E. Of the 57 *Olsenella profusa* sequences detected 19 and 37 were detected in isolate

[3] Novel order in this case only

libraries of patient D and F, respectively, making this a significant difference compared to the molecular libraries of these patients as well as to the other patients ($p < 0.05$). Spirochetes and TM7 were both detected in patient B and D, but the relative occurrence was reversed; i. e. patient B had significantly more TM7 than *Spirochaetes* ($p < 0.05$), while patient D had significantly more of the latter and only few TM7 ($p < 0.05$).

It can be seen in Figure 3.1 that all clonal libraries significantly under-reported *Actinobacteria* numbers compared to culture analysis ($p < <0.05$). Primer 27F YM detected 10 taxa of the phylum *Actinobacteria* and 12 taxa of the phylum *Proteobacteria*, while primer 27F CM discovered only 6 and 10 taxa, respectively.

The detection of *Bacteroidetes* seemed to be influenced mostly by the choice of reverse primer as libraries 3 and 5 that used primer 1387R had a significantly higher rate of detection of *Bacteroidetes* ($p < 0.05$) than the libraries/culture analysis using primer 1492R. The highest detection of *Firmicutes* was seen in the libraries where 27F CM was used.

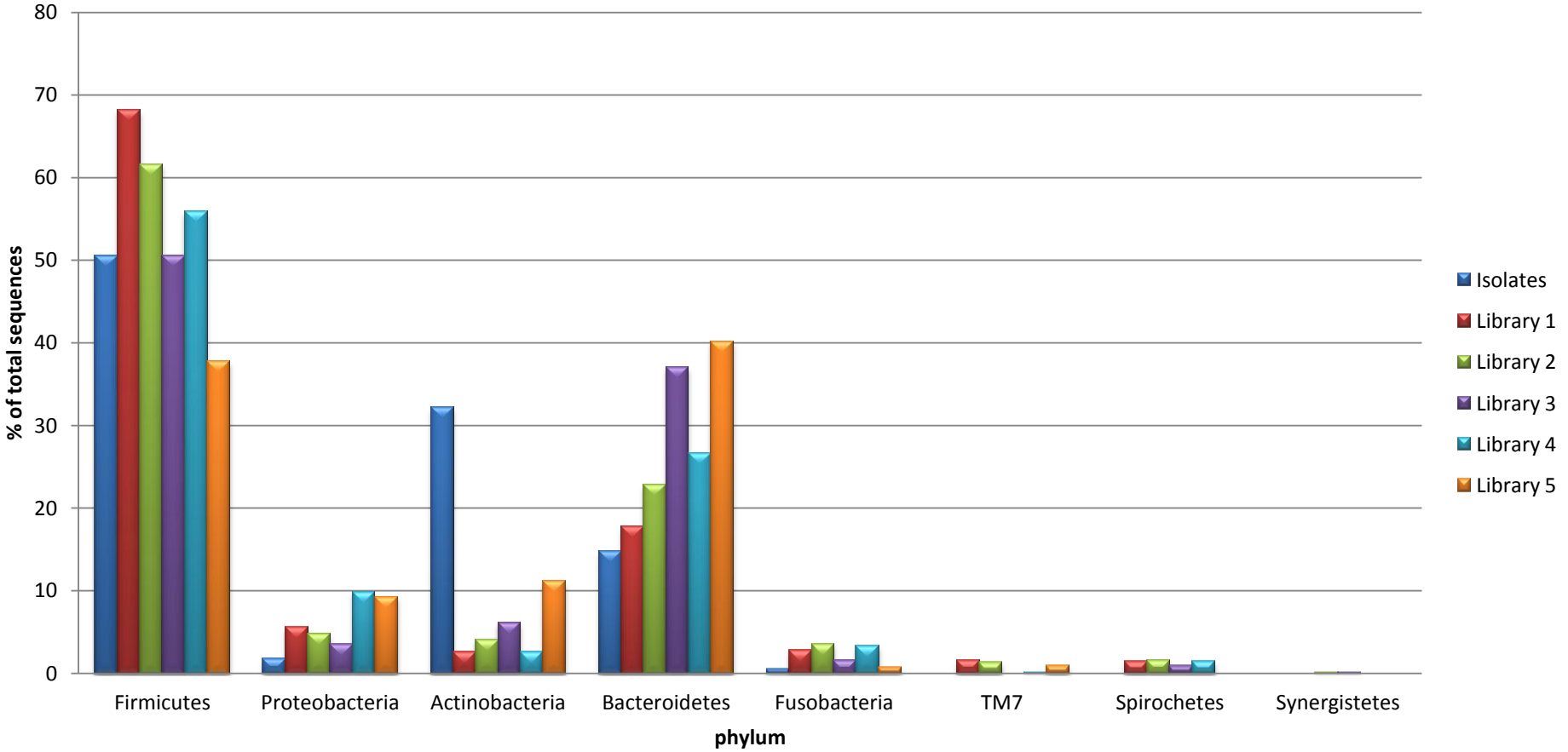


Figure 3.1: Distribution of sequences among bacterial phyla. Primers used: Isolates - 27F CM / 1492R, Library 1 – 27F CM / 1492R, Library 2 – 27F YM / 1492R, Library 3 – 39F / 1387R, Library 4 - 39F / 1492R, Library 5 – 61F / 1387R.

The proportion of the microbiota represented by the genus *Streptococcus* was significantly higher in four of the libraries, compared to culture ($p < 0.05$)(Figure 3.2). In Library 5 levels of streptococci were not significantly lower than those revealed by culture ($p = 0.42372$), but there was a significant difference compared to the other molecular libraries ($p < 0.05$).

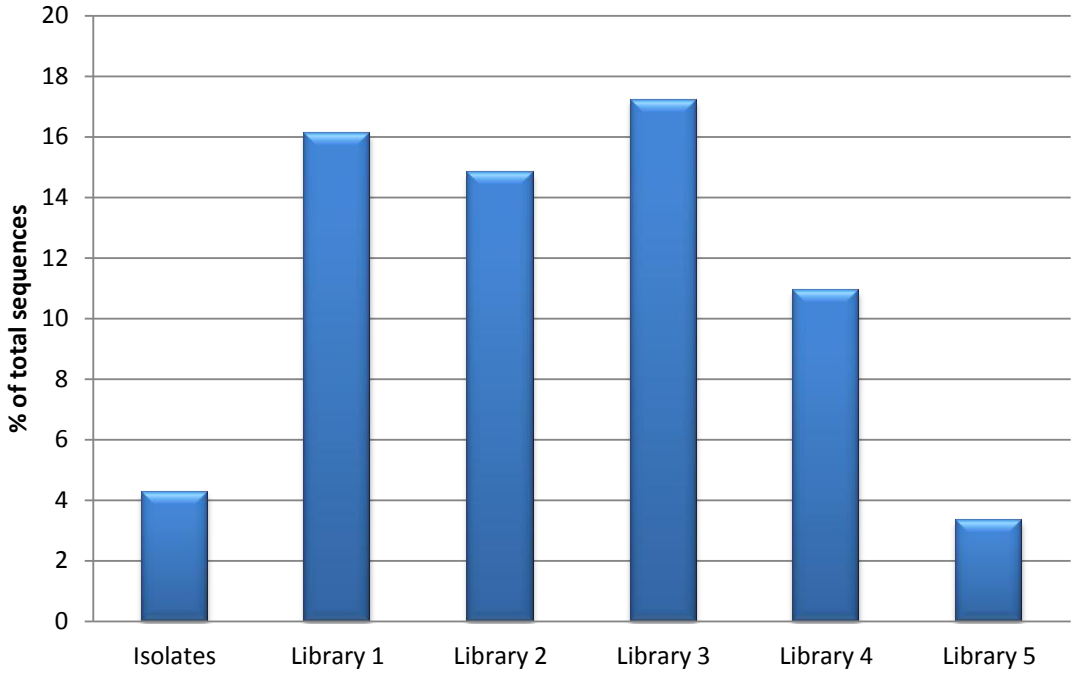


Figure 3.2: Distribution of streptococcal sequences among Isolates and Libraries 1 –5. Primers used: Isolates - 27F CM / 1492R, Library 1 – 27F CM / 1492R, Library 2 – 27F YM / 1492R, Library 3 – 39F / 1387R, Library 4 - 39F / 1492R, Library 5 – 61F / 1387R.

Table 3.5 shows the proportions of selected taxa among the isolates and clones. The taxa shown were chosen because the total number of sequences detected for that species was greater than 1 % of the total and at least one library showed an at least 50 % higher incidence compared to the other libraries or detection rates in two libraries was at most 50 % of the detection rate of the other libraries. For example, libraries prepared with reverse primer 1387R (L3 and L5) detected between 50-75 % less *Pseudoramibacter alactolyticus* compared to reverse primer 1492R ($p < 0.05$). Detection of *Atopobium* OT 416, *Aggregatibacter segnis* and *Haemophilus parainfluenzae* was significantly increased with primer pair 61F / 1387R (L5) to more than double of that detected with culture analysis or any other primer pair combination used ($p < 0.05$). *Olsenella profusa* was only found using culture methods

($p < 0.05$), as were a few other species, but at much lower frequencies (not shown). Detection of *Atopobium rimae, Lactobacillus rhamnosus* and *Propionibacterium* OT 191 showed a similar trend in that detection using culture was on average 11.04, 5.45 and 34.10 times (and therefore significantly, $p < 0.05$) higher than that using molecular methods, respectively. Detection of *Fusobacteria* was significantly reduced when culture methods or reverse primer 1387R (Lib 3 and Lib 5) were used; detection of *Prevotella oralis* and *Prevotella tannerae* was significantly increased using primer pair 39F / 1387R (L3) ($p < 0.05$).

Table 3.5: Detection of selected taxa by culture and molecular analysis

		% of total				
culture/library	Isolates	Lib 1	Lib 2	Lib 3	Lib 4	Lib 5
Species						
***Pseudoramibacter alactolyticus* OT 538**	2.96	3.52	3.52	1.48	4.63	0.93
***Atopobium* OT 416**	1.85	0.93	0.56	0.93	0.37	4.26
Aggregatibacter segnis	0.00	1.11	0.74	0.74	1.67	3.33
Haemophilus parainfluenzae	0.56	1.30	0.93	0.56	0.93	3.33
Olsenella profusa	10.56	0.00	0.00	0.00	0.00	0.00
***Atopobium* rimae**	4.26	0.74	0.37	0.19	0.55	0.55
Lactobacillus rhamnosus	5.74	1.67	0.74	1.30	1.67	1.48
***Propionibacterium* sp. OT 191**	6.48	0.00	0.19	0.00	0.19	0.00
Fusobacterium nucleatum	0.37	2.04	2.41	1.11	2.59	0.56
Prevotella oralis	1.11	0.74	0.93	2.78	0.56	0.56
Prevotella tannerae	0.74	1.67	3.15	7.59	3.52	4.26

3.4.2 Phylogenetic trees

Phylogenetic trees of all species/sequences detected in the patient samples were constructed (Figure 3.3 - Figure 3.10), together with trees showing the position of the novel taxa identified in the study (Figure 3.11 - Figure 3.21). Isolates identified as novel are highlighted in bold script in all trees.

3.4.2.1 Phylogenetic trees of all detected sequences

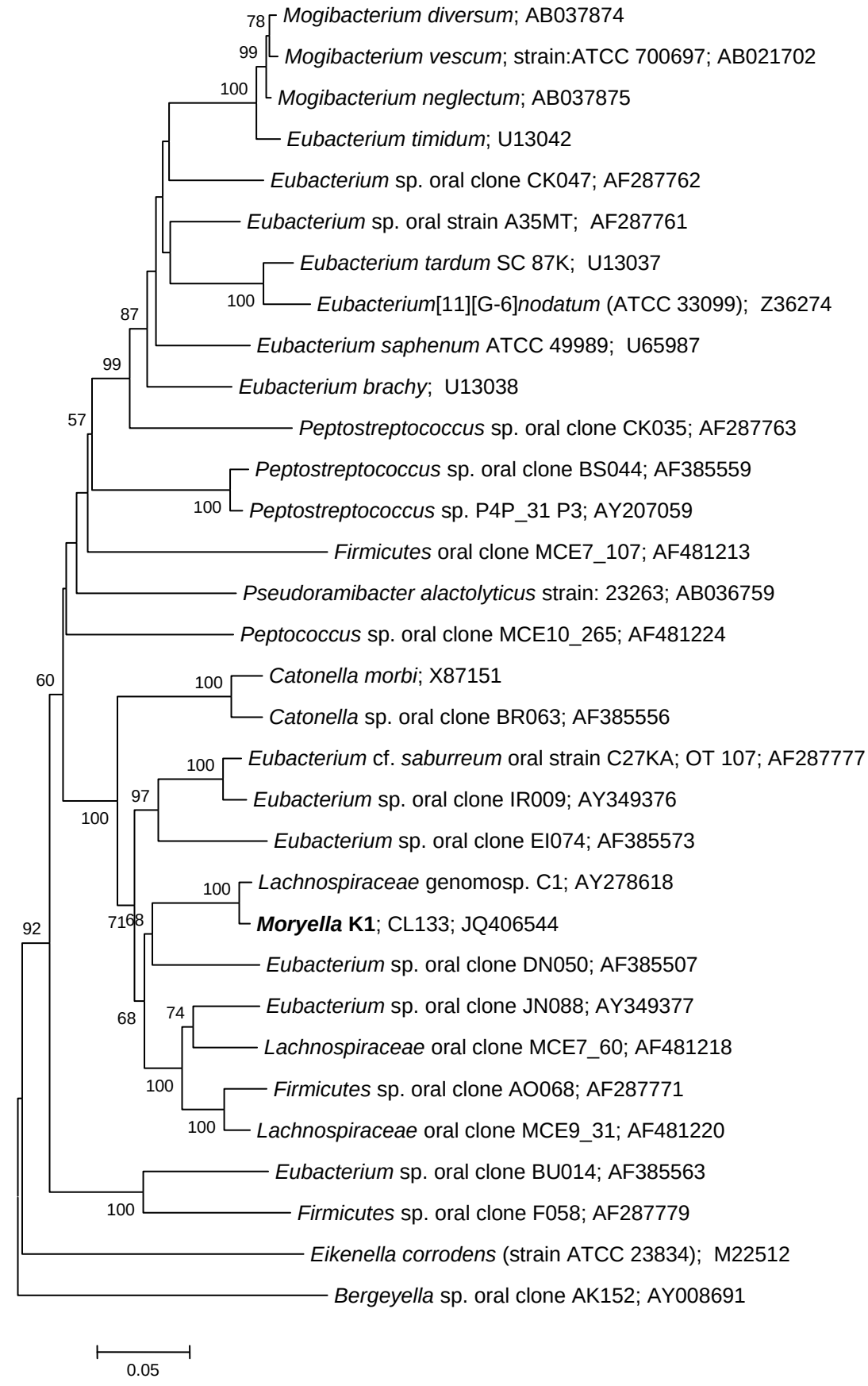


Figure 3.3: Phylogenetic tree based on 16S rRNA gene sequence comparisons over 1218 aligned bases from 32 nucleotide sequences showing the relationship of taxa detected in patient samples belonging to the phylum

***Firmicutes*. The tree was constructed using the neighbour-joining method from a distance matrix constructed from aligned sequences using the Jukes-Cantor correction. Numbers represent bootstrap values for each branch based on data for 1000 trees. Accession numbers for 16S rRNA gene sequences are given for each strain. Scale bar shows number of nucleotide substitutions per site.**

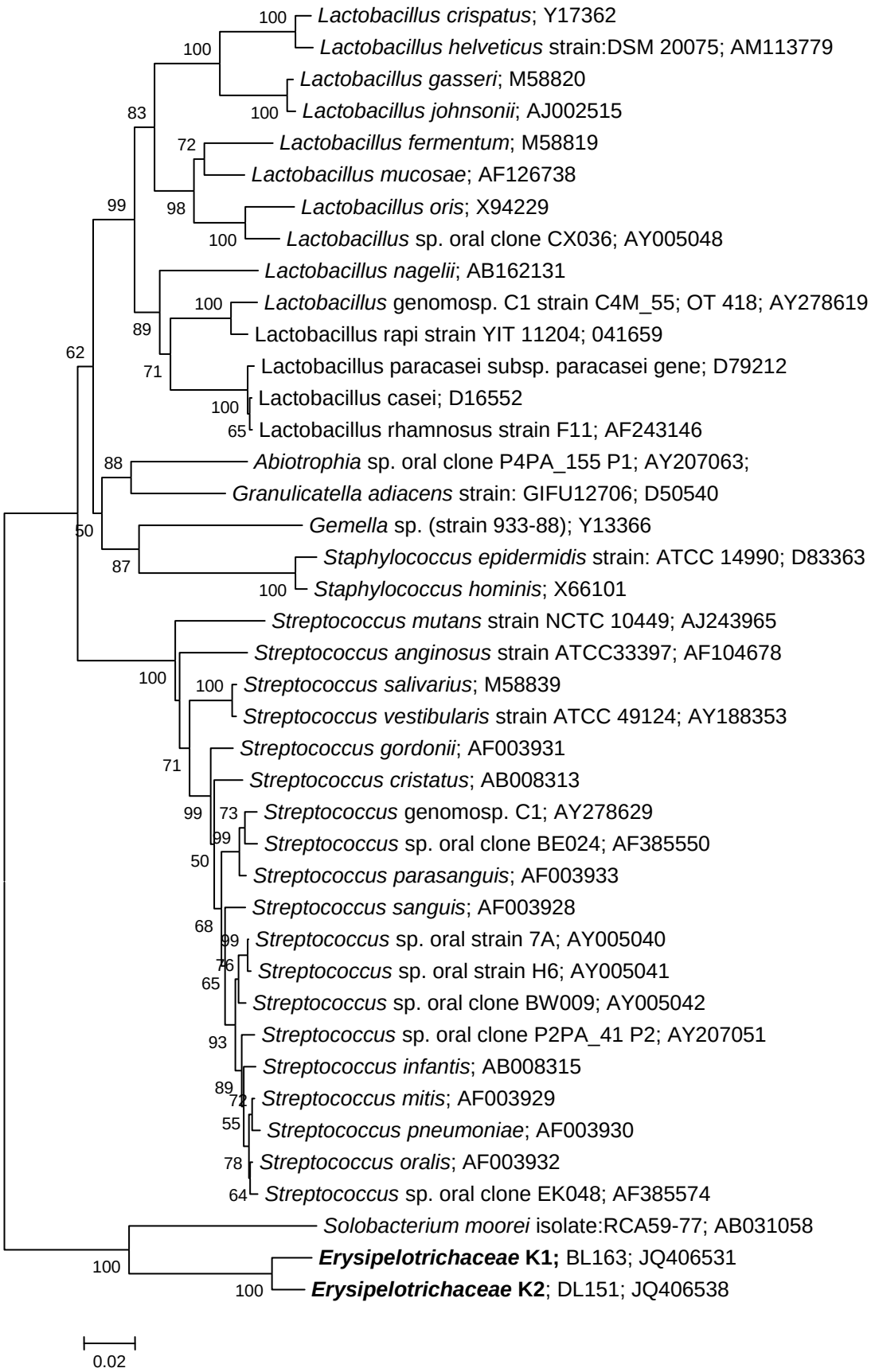


Figure 3.4: Phylogenetic tree based on 16S rRNA gene sequence comparisons over 1248 aligned bases from 41 nucleotide sequences showing the relationship of taxa detected in patient samples belonging to the phylum *Firmicutes*. The tree was constructed using the neighbour-joining method from a distance matrix constructed from aligned sequences using the Jukes-Cantor correction. Numbers represent bootstrap values for

each branch based on data for 1000 trees. Accession numbers for 16S rRNA gene sequences are given for each strain. Scale bar shows number of nucleotide substitutions per site.

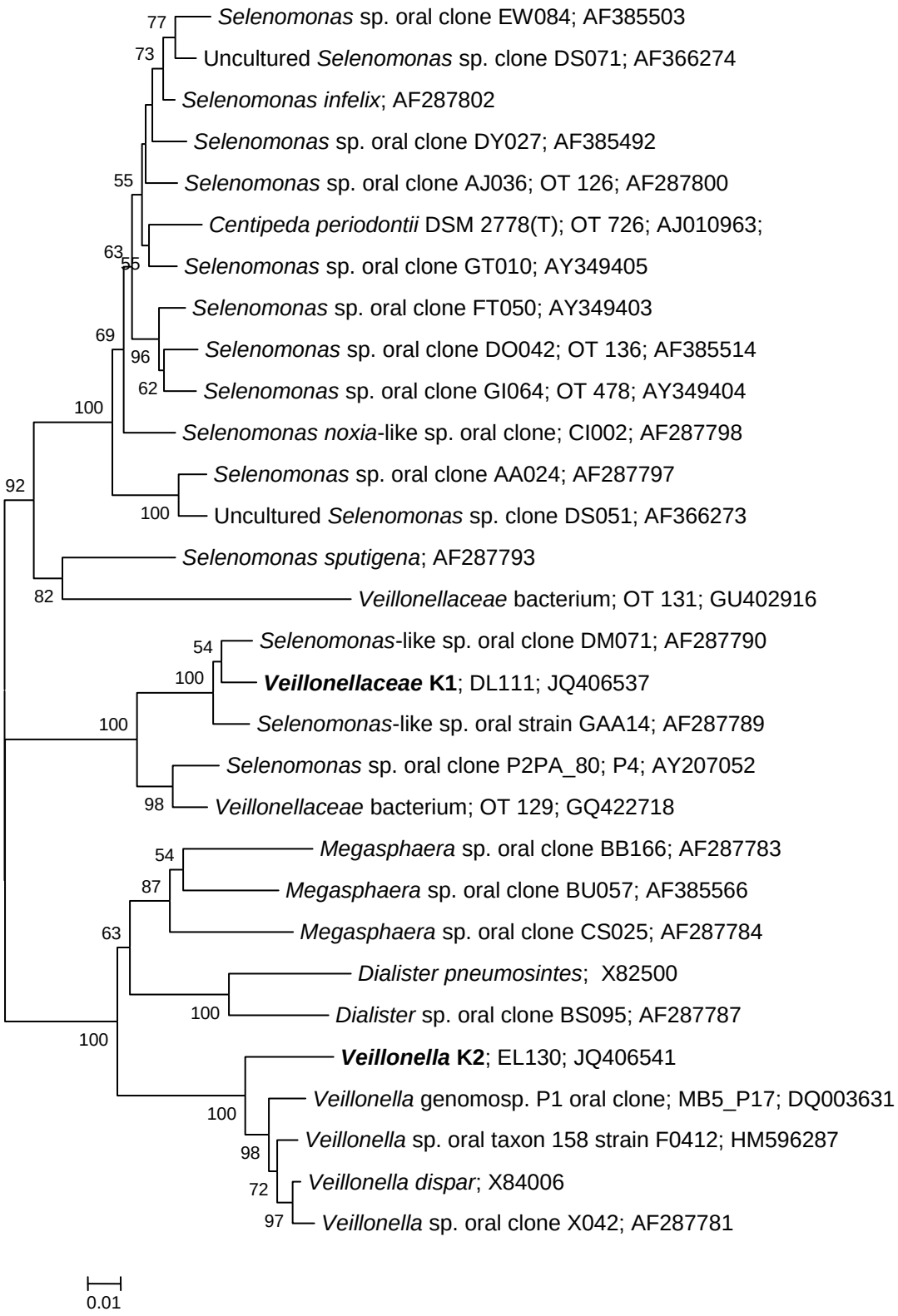


Figure 3.5: Phylogenetic tree based on 16S rRNA gene sequence comparisons over 1113 aligned bases from 30 nucleotide sequences showing the relationship of taxa detected in patient samples belonging to the phylum *Firmicutes*. The tree was constructed using the neighbour-joining method from a distance matrix constructed from aligned sequences using the Jukes-Cantor correction. Numbers represent bootstrap values for each branch based on data for 1000 trees. Accession numbers for 16S rRNA gene sequences are given for each strain. Scale bar shows number of nucleotide substitutions per site.

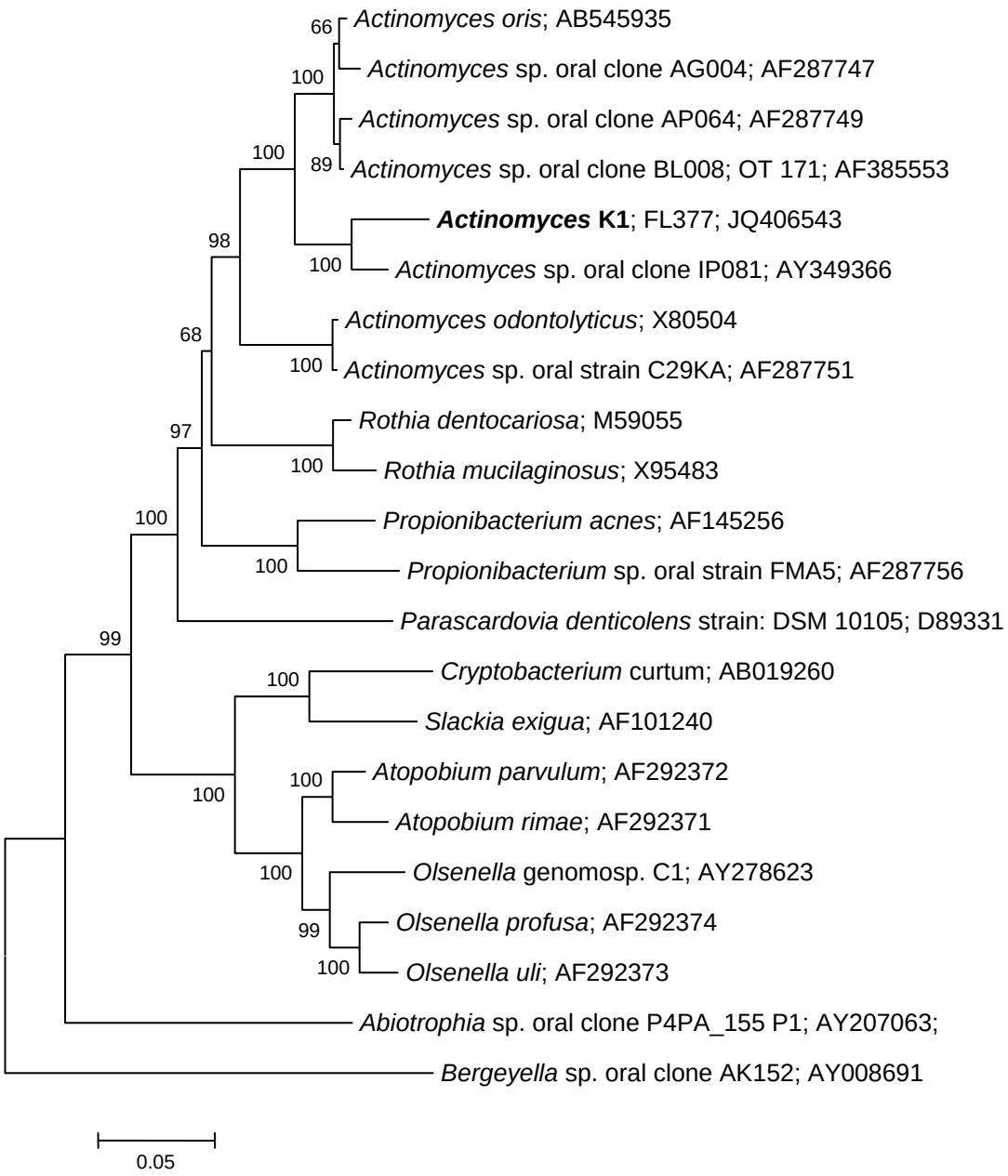


Figure 3.6: Phylogenetic tree based on 16S rRNA gene sequence comparisons over 1252 aligned bases from 22 nucleotide sequences showing the relationship of taxa detected in patient samples belonging to the phylum *Actinobacteria*. The tree was constructed using the neighbour-joining method from a distance matrix constructed from aligned sequences using the Jukes-Cantor correction. Numbers represent bootstrap values for each branch based on data for 1000 trees. Accession numbers for 16S rRNA gene sequences are given for each strain. Scale bar shows number of nucleotide substitutions per site.

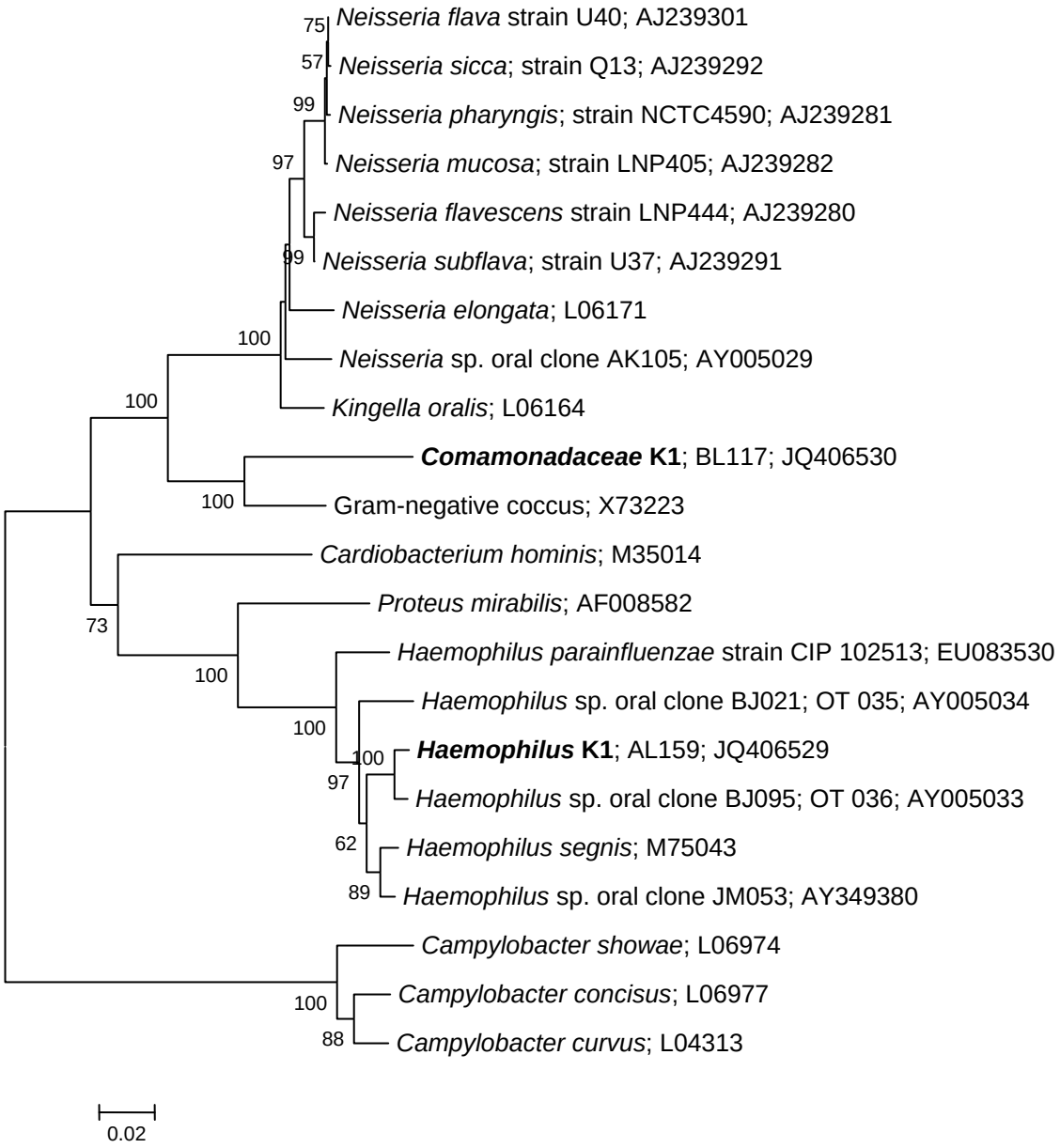


Figure 3.7: Phylogenetic tree based on 16S rRNA gene sequence comparisons over 1201 aligned bases from 22 nucleotide sequences showing the relationship of taxa detected in patient samples belonging to the phylum *Proteobacteria*. The tree was constructed using the neighbour-joining method from a distance matrix constructed from aligned sequences using the Jukes-Cantor correction. Numbers represent bootstrap values for each branch based on data for 1000 trees. Accession numbers for 16S rRNA gene sequences are given for each strain. Scale bar shows number of nucleotide substitutions per site.

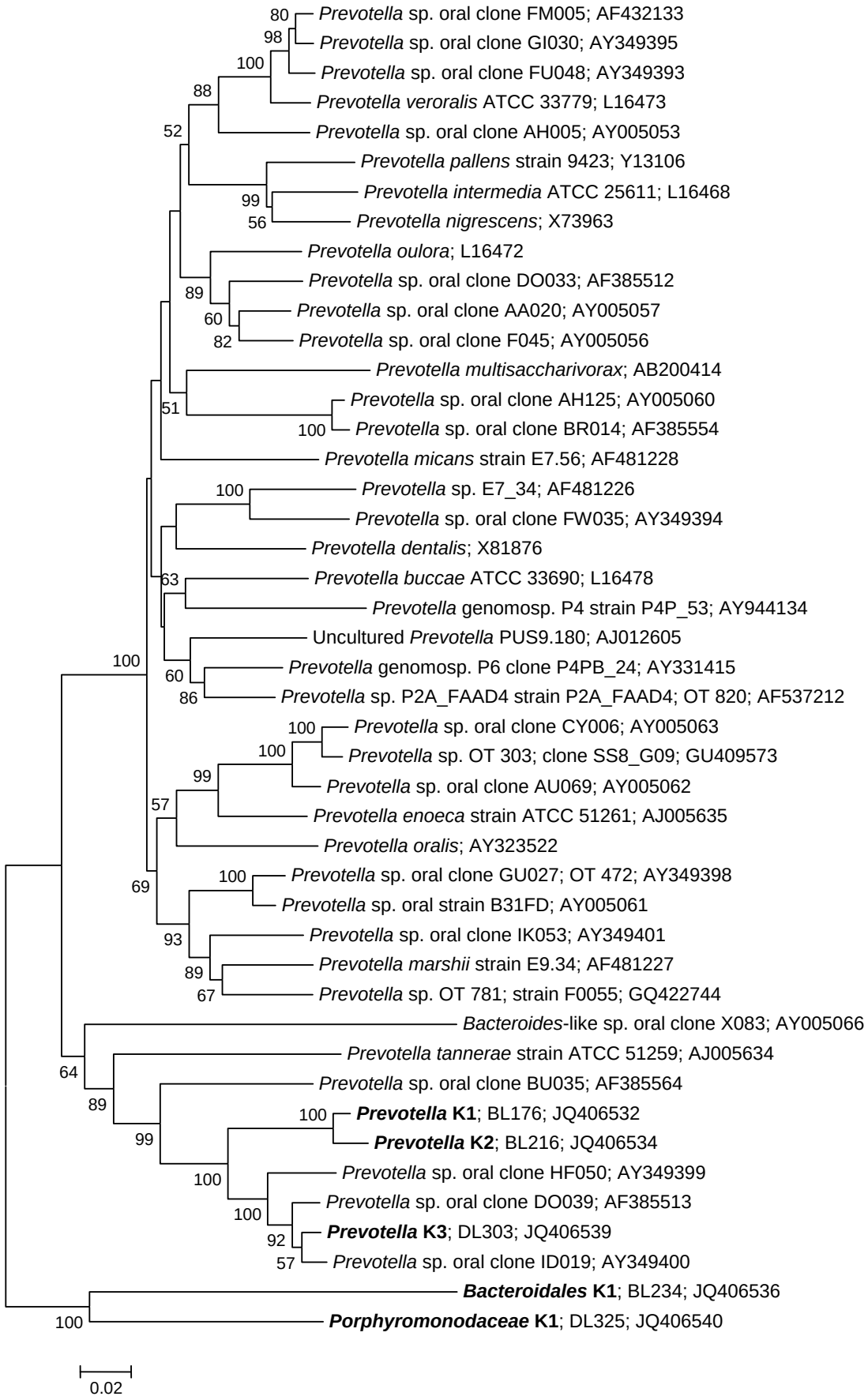


Figure 3.8: Phylogenetic tree based on 16S rRNA gene sequence comparisons over 942 aligned bases from 45 nucleotide sequences showing the relationship of taxa detected in patient samples belonging to the phylum

***Bacteroidetes*. The tree was constructed using the neighbour-joining method from a distance matrix constructed from aligned sequences using the Jukes-Cantor correction. Numbers represent bootstrap values for each branch based on data for 1000 trees. Accession numbers for 16S rRNA gene sequences are given for each strain. Scale bar shows number of nucleotide substitutions per site.**

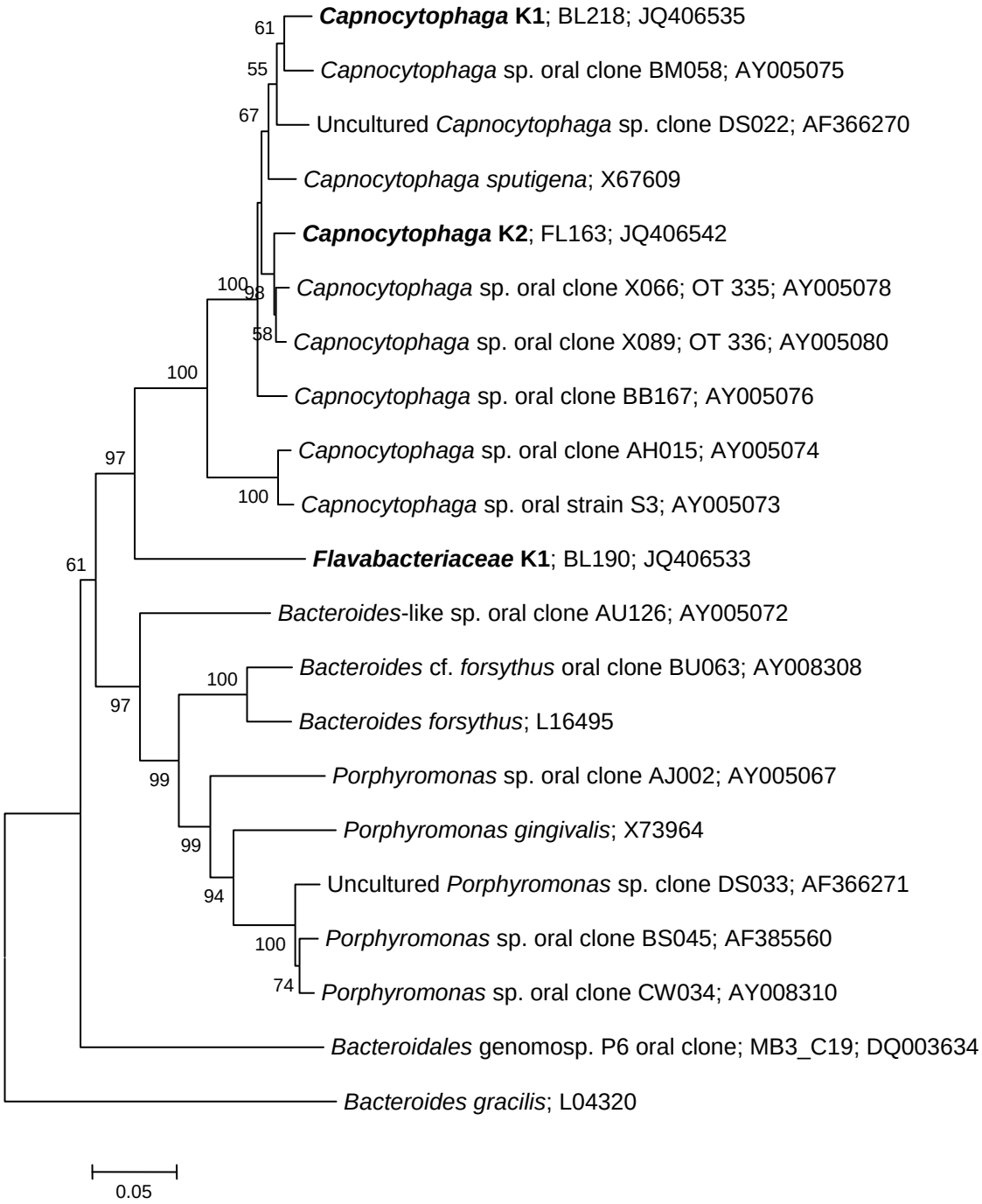


Figure 3.9: Phylogenetic tree based on 16S rRNA gene sequence comparisons over 1042 aligned bases from 21 nucleotide sequences showing the relationship of taxa detected in patient samples belonging to the phylum *Bacteroidetes*. The tree was constructed using the neighbour-joining method from a distance matrix constructed from aligned sequences using the Jukes-Cantor correction. Numbers represent bootstrap values for each branch based on data for 1000 trees. Accession numbers for 16S rRNA gene sequences are given for each strain. Scale bar shows number of nucleotide substitutions per site.

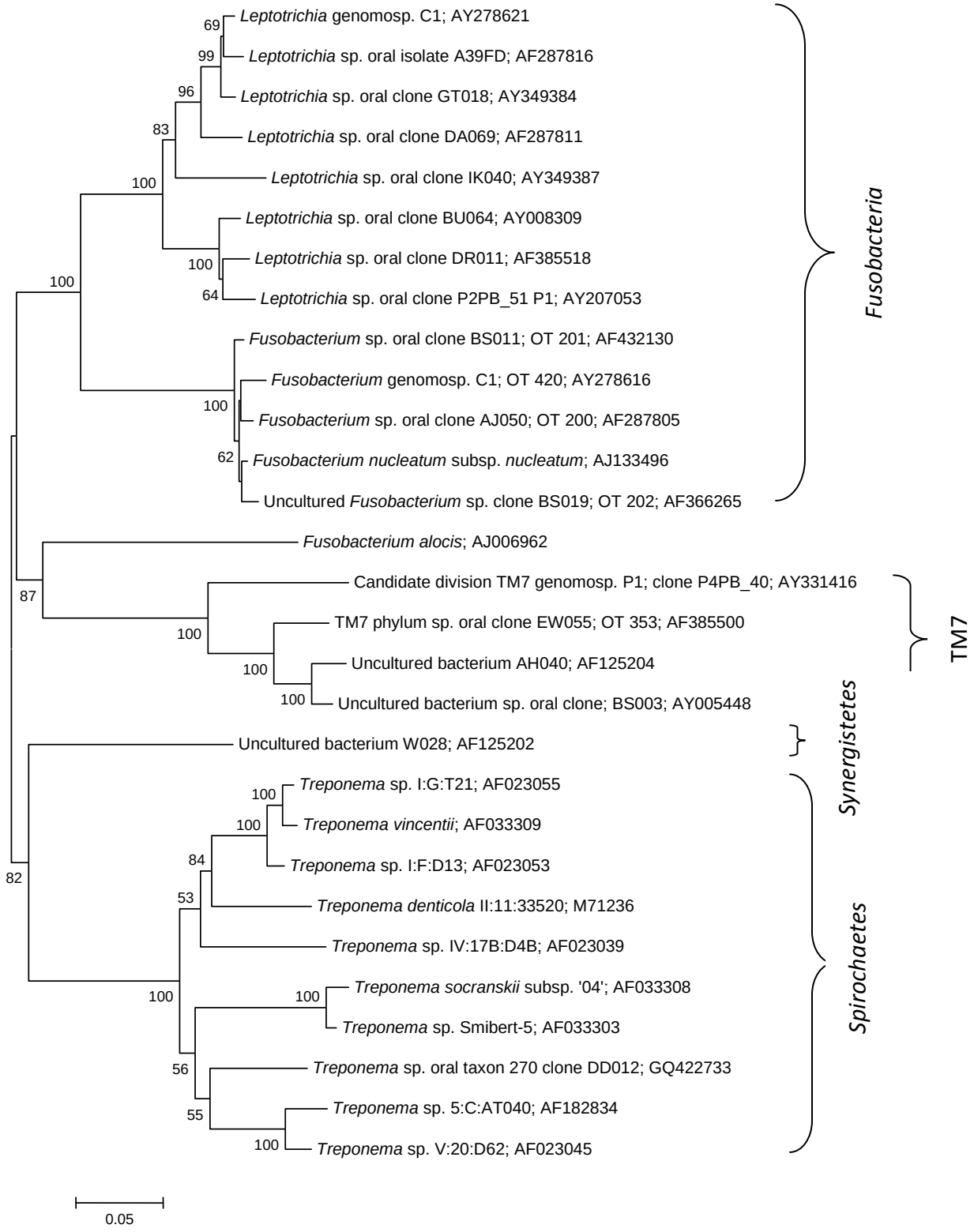


Figure 3.10: Phylogenetic tree based on 16S rRNA gene sequence comparisons over 1029 aligned bases from 29 nucleotide sequences showing the relationship of taxa detected in patient samples belonging to the phyla *Spirochaetes, Synergistetes, TM7* and *Fusobacteria*. The tree was constructed using the neighbour-joining method from a distance matrix constructed from aligned sequences using the Jukes-Cantor correction. Numbers represent bootstrap values for each branch based on data for 1000 trees. Accession numbers for 16S rRNA gene sequences are given for each strain. Scale bar shows number of nucleotide substitutions per site.

3.4.2.2 Phylogenetic trees of novel sequences

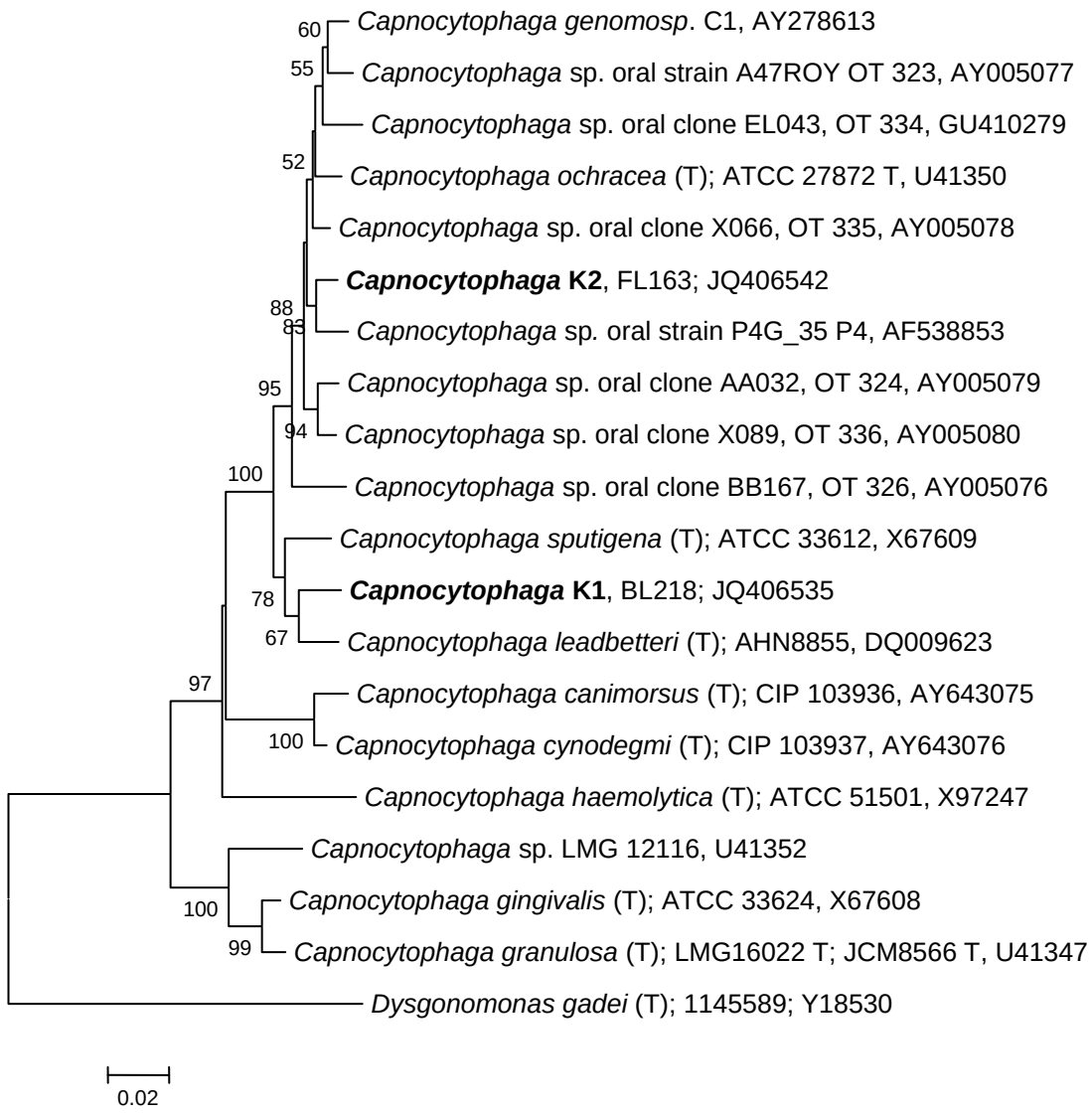


Figure 3.11: Phylogenetic tree based on 16S rRNA gene sequence comparisons over 1332 aligned bases from 20 nucleotide sequences showing the relationship between *Capnocytophaga* K1 and K2 and the genus *Capnocytophaga*. The tree was constructed using the neighbour-joining method from a distance matrix constructed from aligned sequences using the Jukes-Cantor correction. Numbers represent bootstrap values for each branch based on data for 1000 trees. Accession numbers for 16S rRNA gene sequences are given for each strain. Scale bar shows number of nucleotide substitutions per site.

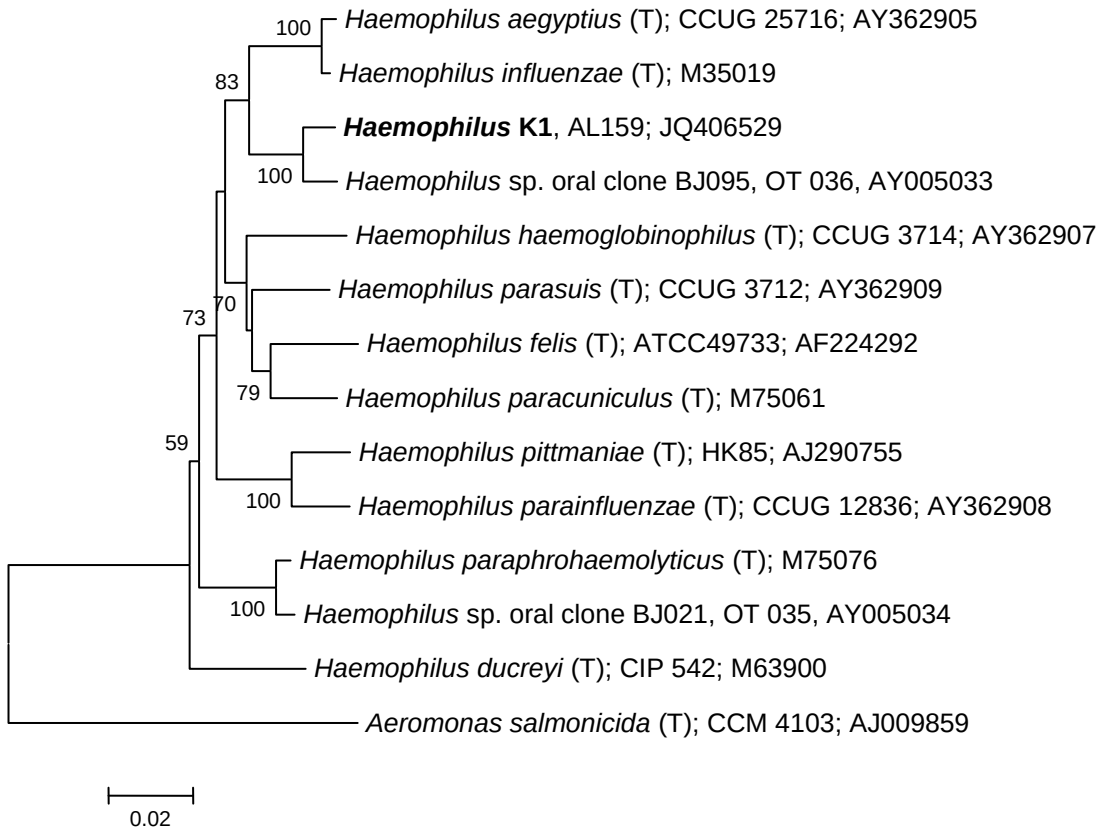


Figure 3.12: Phylogenetic tree based on 16S rRNA gene sequence comparisons over 1293 aligned bases from 14 nucleotide sequences showing the relationship between *Haemophilus* K1 and the genus *Haemophilus*. The tree was constructed using the neighbour-joining method from a distance matrix constructed from aligned sequences using the Jukes-Cantor correction. Numbers represent bootstrap values for each branch based on data for 1000 trees. Accession numbers for 16S rRNA gene sequences are given for each strain. Scale bar shows number of nucleotide substitutions per site.

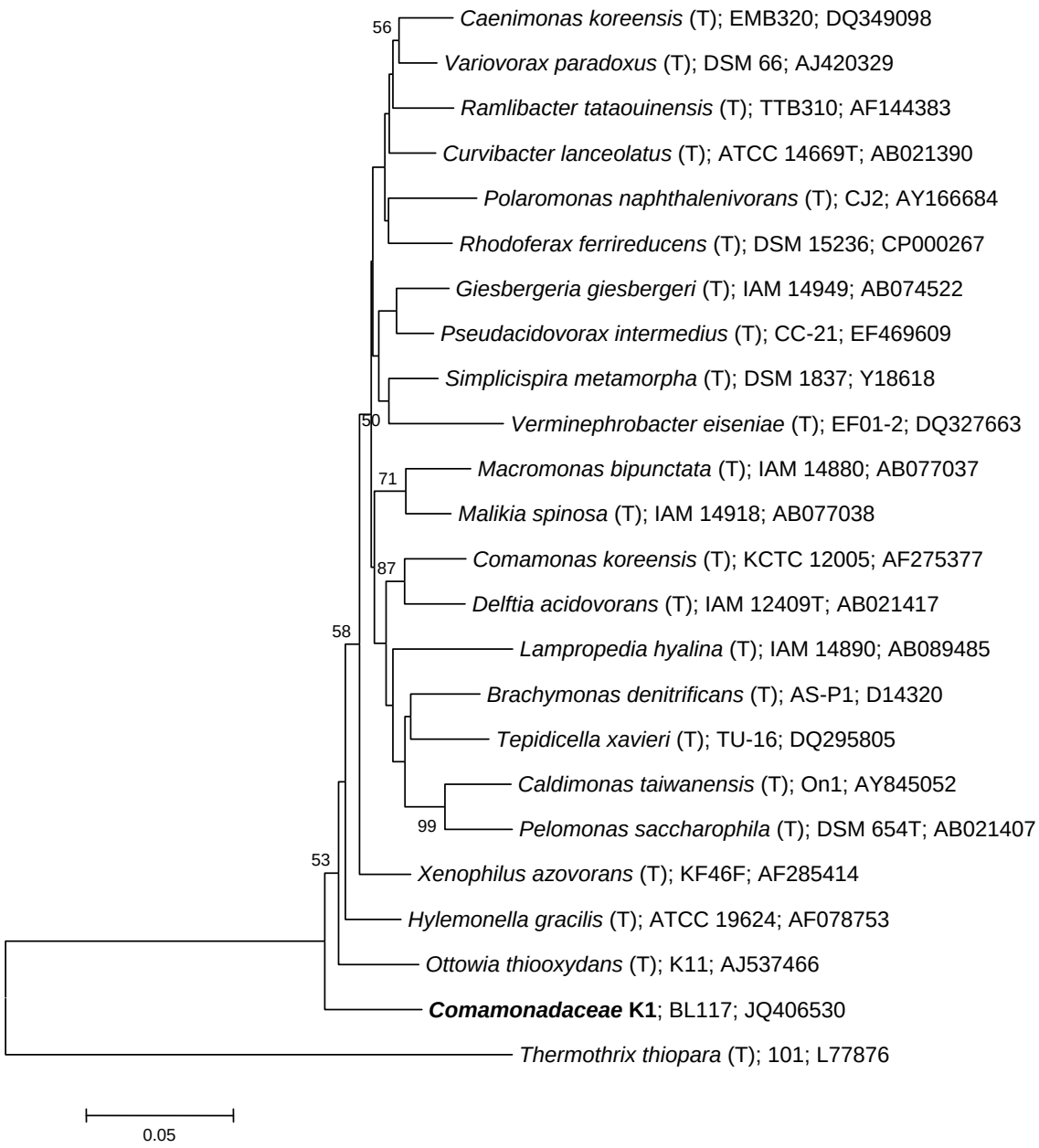


Figure 3.13: Phylogenetic tree based on 16S rRNA gene sequence comparisons over 1343 aligned bases from 24 nucleotide sequences showing the relationship between *Comamonadaceae* K1 and representatives from the family *Comamonadaceae*. The tree was constructed using the neighbour-joining method from a distance matrix constructed from aligned sequences using the Jukes-Cantor correction. Numbers represent bootstrap values for each branch based on data for 1000 trees. Accession numbers for 16S rRNA gene sequences are given for each strain. Scale bar shows number of nucleotide substitutions per site.

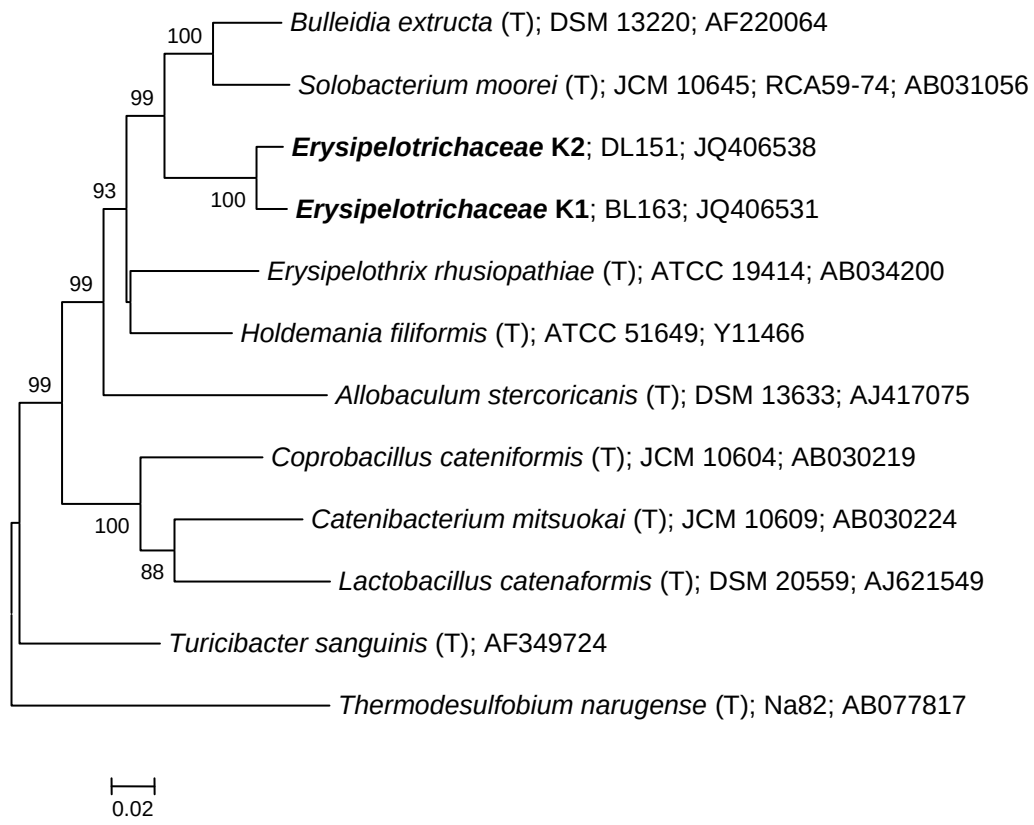


Figure 3.14: Phylogenetic tree based on 16S rRNA gene sequence comparisons over 1231 aligned bases from 12 nucleotide sequences showing the relationship between *Erysipelotrichaceae* K1 and K2 and representatives from the family *Erysipelotrichaceae*. The tree was constructed using the neighbour-joining method from a distance matrix constructed from aligned sequences using the Jukes-Cantor correction. Numbers represent bootstrap values for each branch based on data for 1000 trees. Accession numbers for 16S rRNA gene sequences are given for each strain. Scale bar shows number of nucleotide substitutions per site.

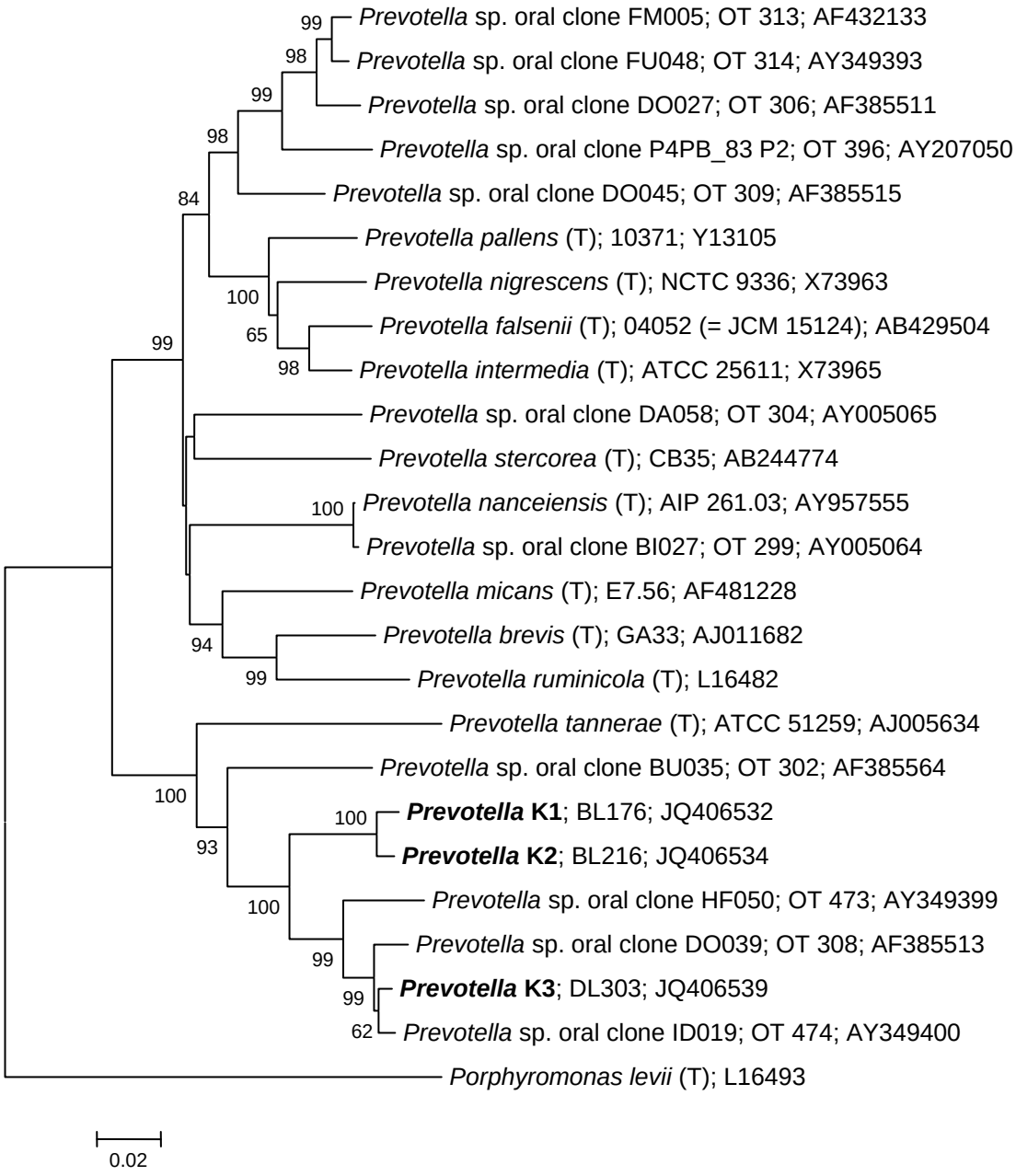


Figure 3.15: Phylogenetic tree based on 16S rRNA gene sequence comparisons over 968 aligned bases from 25 nucleotide sequences showing the relationship between *Prevotella* K1 to K3 and representatives from the genus *Prevotella*. The tree was constructed using the neighbour-joining method from a distance matrix constructed from aligned sequences using the Jukes-Cantor correction. Numbers represent bootstrap values for each branch based on data for 1000 trees. Accession numbers for 16S rRNA gene sequences are given for each strain. Scale bar shows number of nucleotide substitutions per site.

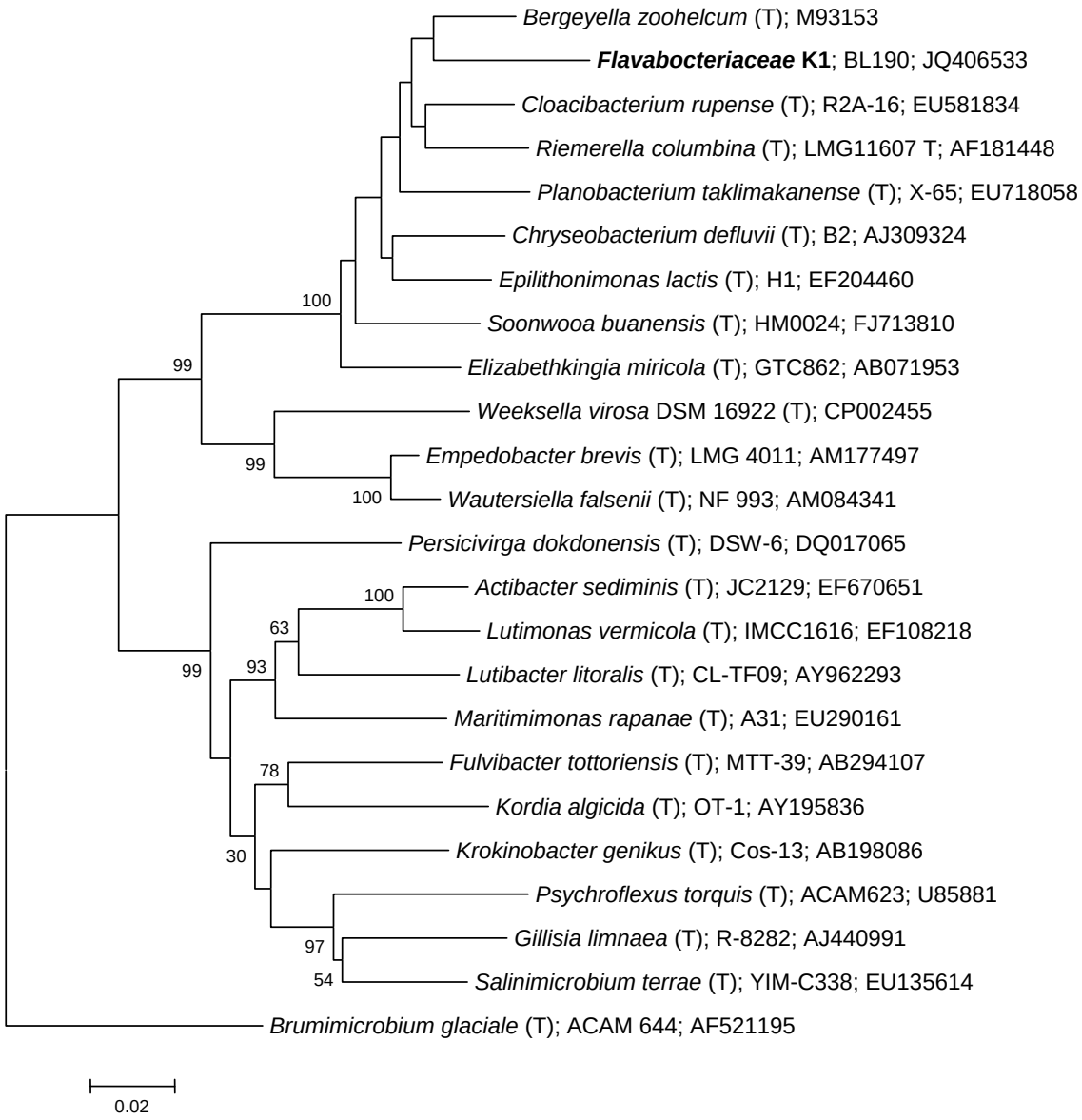


Figure 3.16: Phylogenetic tree based on 16S rRNA gene sequence comparisons over 1159 aligned bases from 24 nucleotide sequences showing the relationship between *Flavabocteriaceae* K1 and representatives from the family *Flavabocteriaceae*. The tree was constructed using the neighbour-joining method from a distance matrix constructed from aligned sequences using the Jukes-Cantor correction. Numbers represent bootstrap values for each branch based on data for 1000 trees. Accession numbers for 16S rRNA gene sequences are given for each strain. Scale bar shows number of nucleotide substitutions per site.

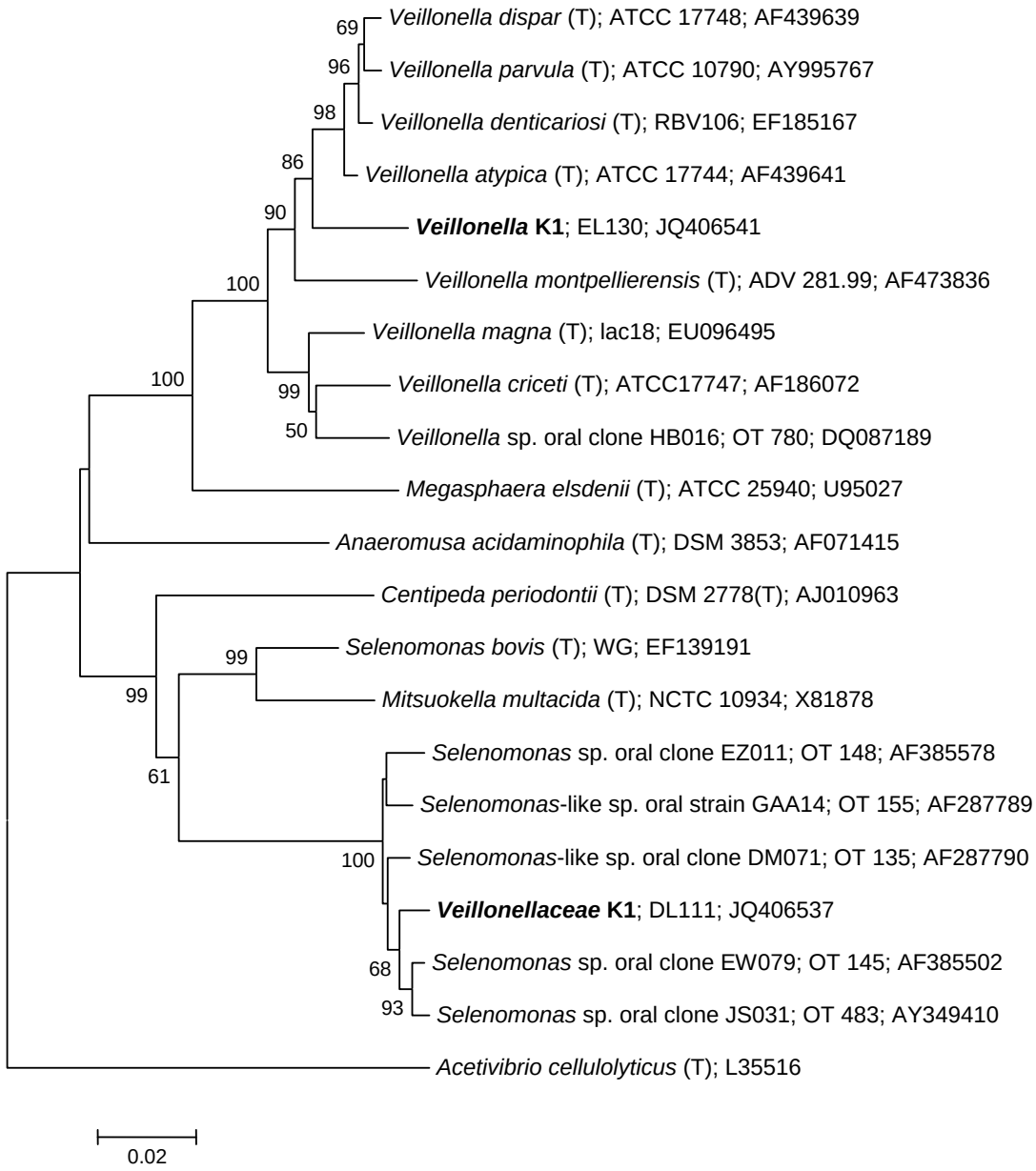


Figure 3.17: Phylogenetic tree based on 16S rRNA gene sequence comparisons over 1162 aligned bases from 21 nucleotide sequences showing the relationship between *Veillonella* K1 and *Veillonellaceae* K1 and representatives from the family *Veillonellaceae*. The tree was constructed using the neighbour-joining method from a distance matrix constructed from aligned sequences using the Jukes-Cantor correction. Numbers represent bootstrap values for each branch based on data for 1000 trees. Accession numbers for 16S rRNA gene sequences are given for each strain. Scale bar shows number of nucleotide substitutions per site.

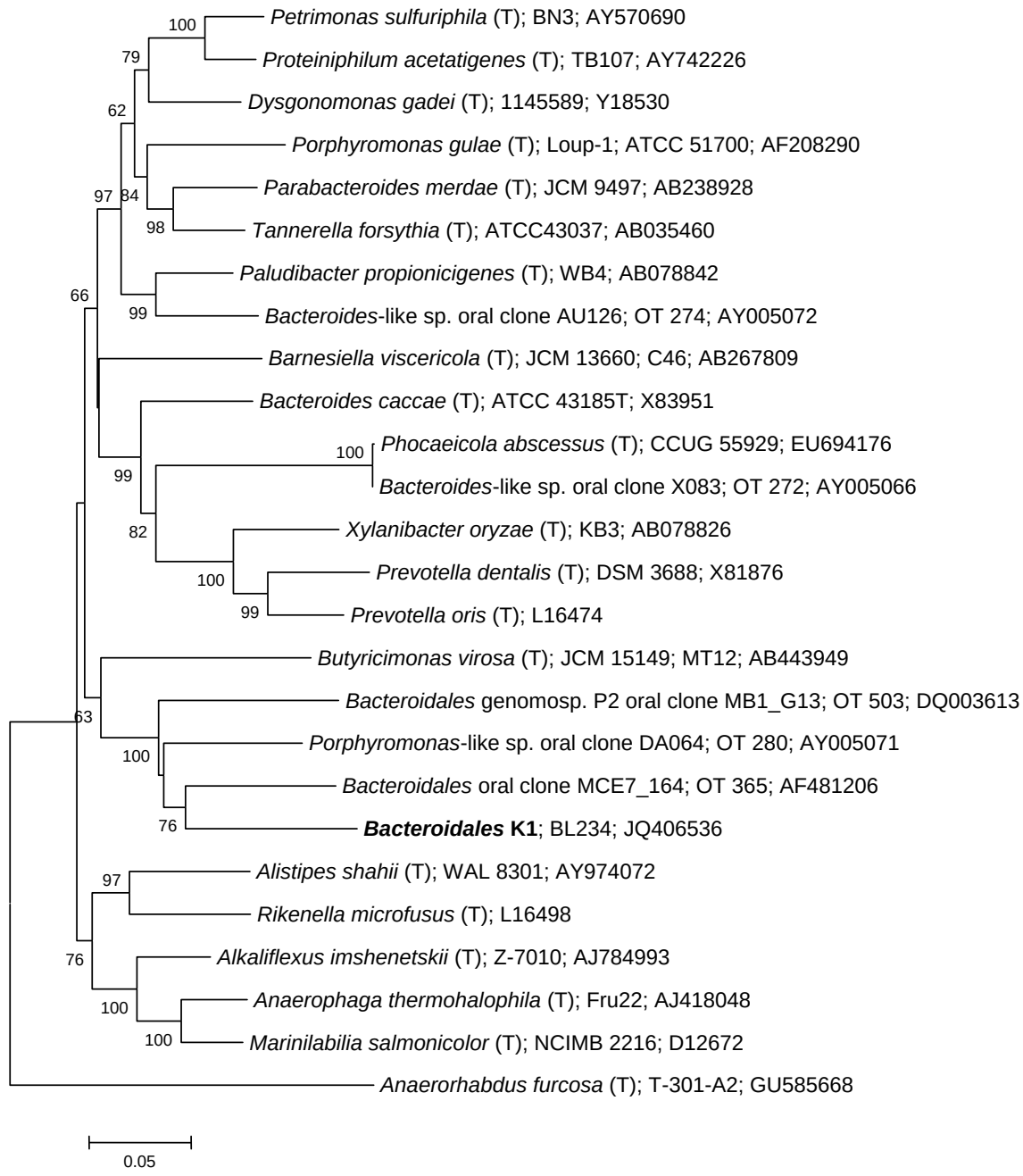


Figure 3.18: Phylogenetic tree based on 16S rRNA gene sequence comparisons over 1130 aligned bases from 26 nucleotide sequences showing the relationship between *Bacteroidales* K1 and representatives from the phylum *Bacteroidetes*. The tree was constructed using the neighbour-joining method from a distance matrix constructed from aligned sequences using the Jukes-Cantor correction. Numbers represent bootstrap values for each branch based on data for 1000 trees. Accession numbers for 16S rRNA gene sequences are given for each strain. Scale bar shows number of nucleotide substitutions per site.

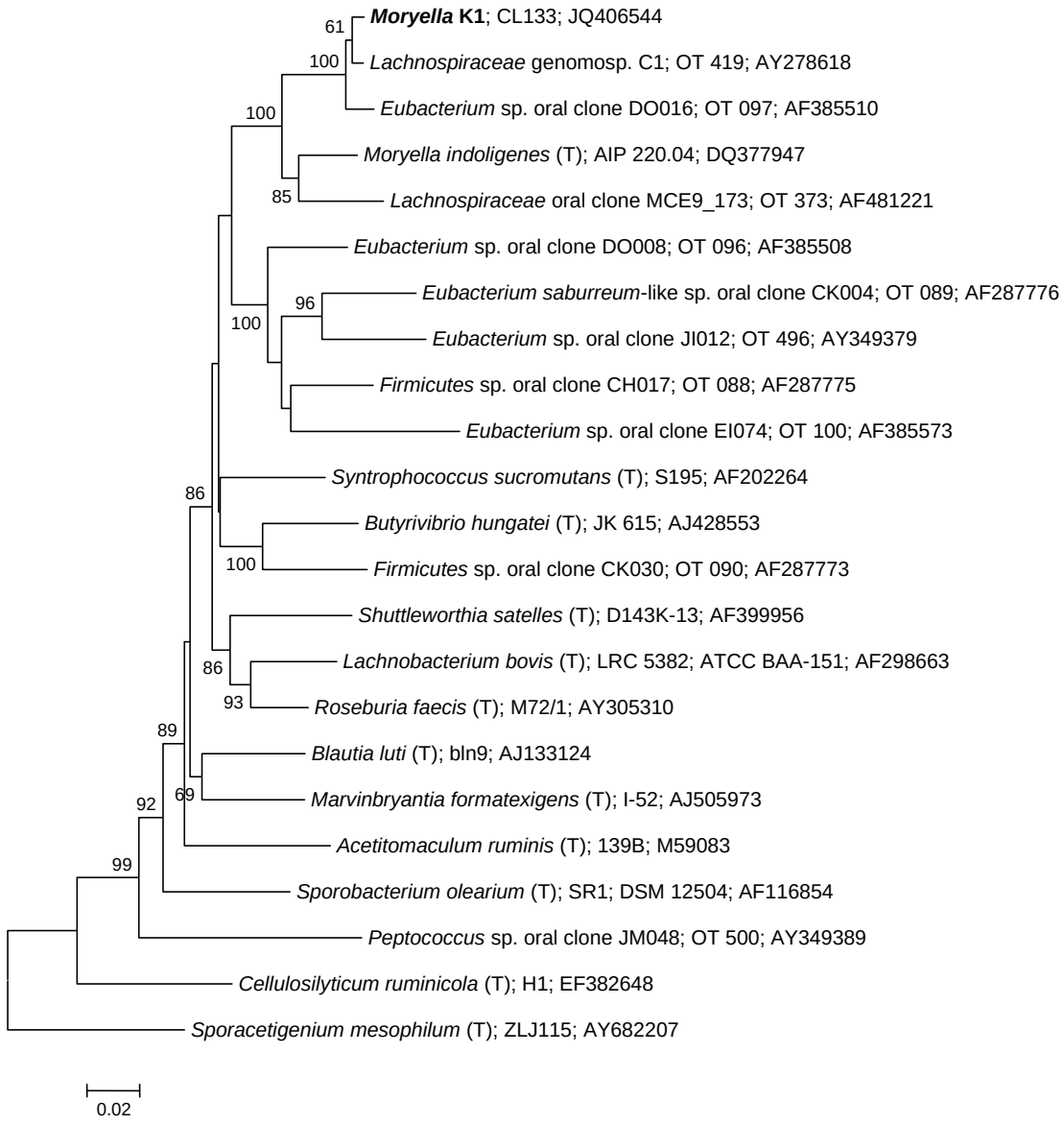


Figure 3.19: Phylogenetic tree based on 16S rRNA gene sequence comparisons over 1165 aligned bases from 23 nucleotide sequences showing the relationship between *Moryella* K1 and representatives from the family *Lachnospiraceae*. The tree was constructed using the neighbour-joining method from a distance matrix constructed from aligned sequences using the Jukes-Cantor correction. Numbers represent bootstrap values for each branch based on data for 1000 trees. Accession numbers for 16S rRNA gene sequences are given for each strain. Scale bar shows number of nucleotide substitutions per site.

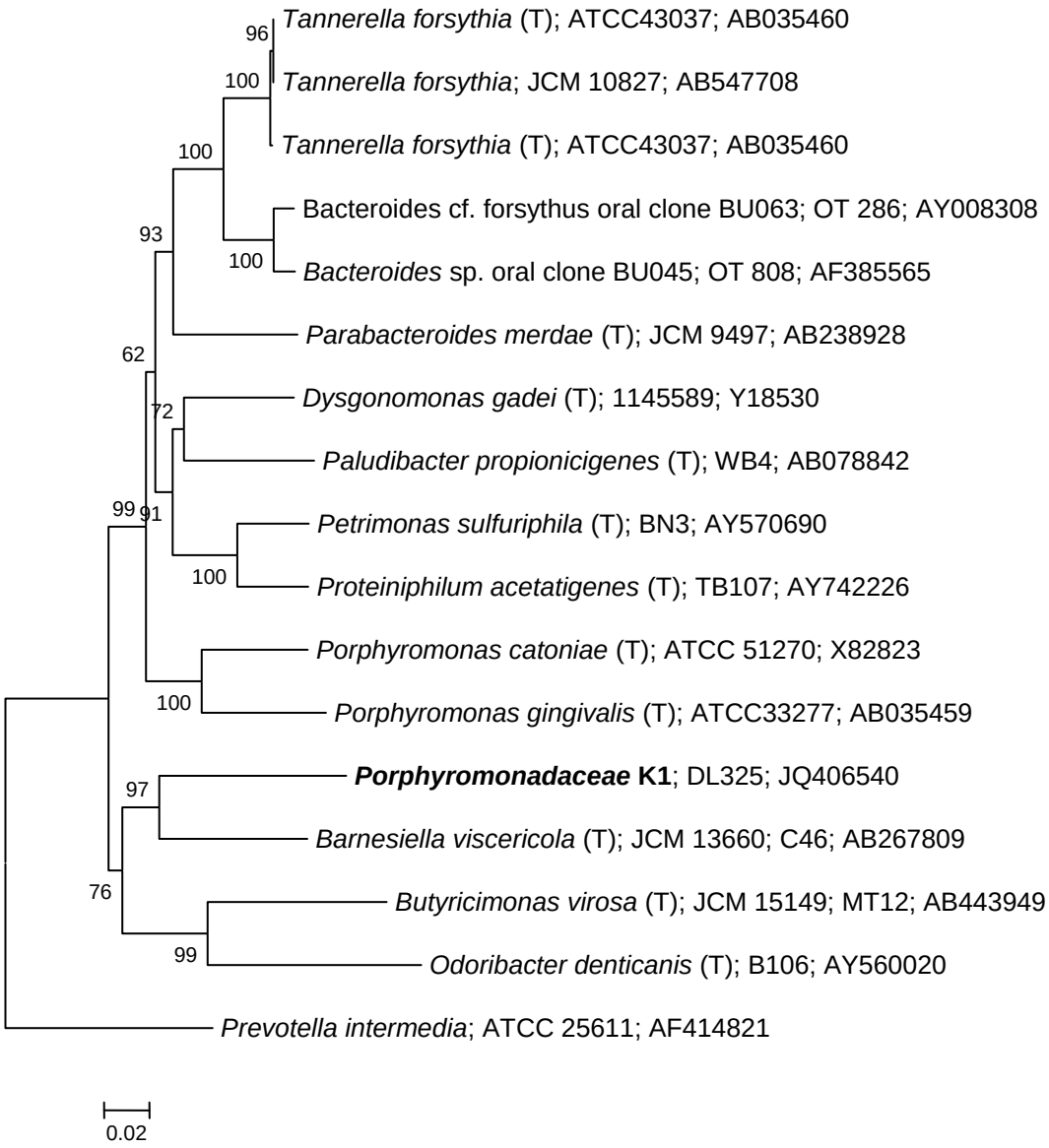


Figure 3.20: Phylogenetic tree based on 16S rRNA gene sequence comparisons over 1275 aligned bases from 17 nucleotide sequences showing the relationship between *Porphyromonadaceae* K1 and representatives from the family *Porphyromonadaceae*. The tree was constructed using the neighbour-joining method from a distance matrix constructed from aligned sequences using the Jukes-Cantor correction. Numbers represent bootstrap values for each branch based on data for 1000 trees. Accession numbers for 16S rRNA gene sequences are given for each strain. Scale bar shows number of nucleotide substitutions per site.

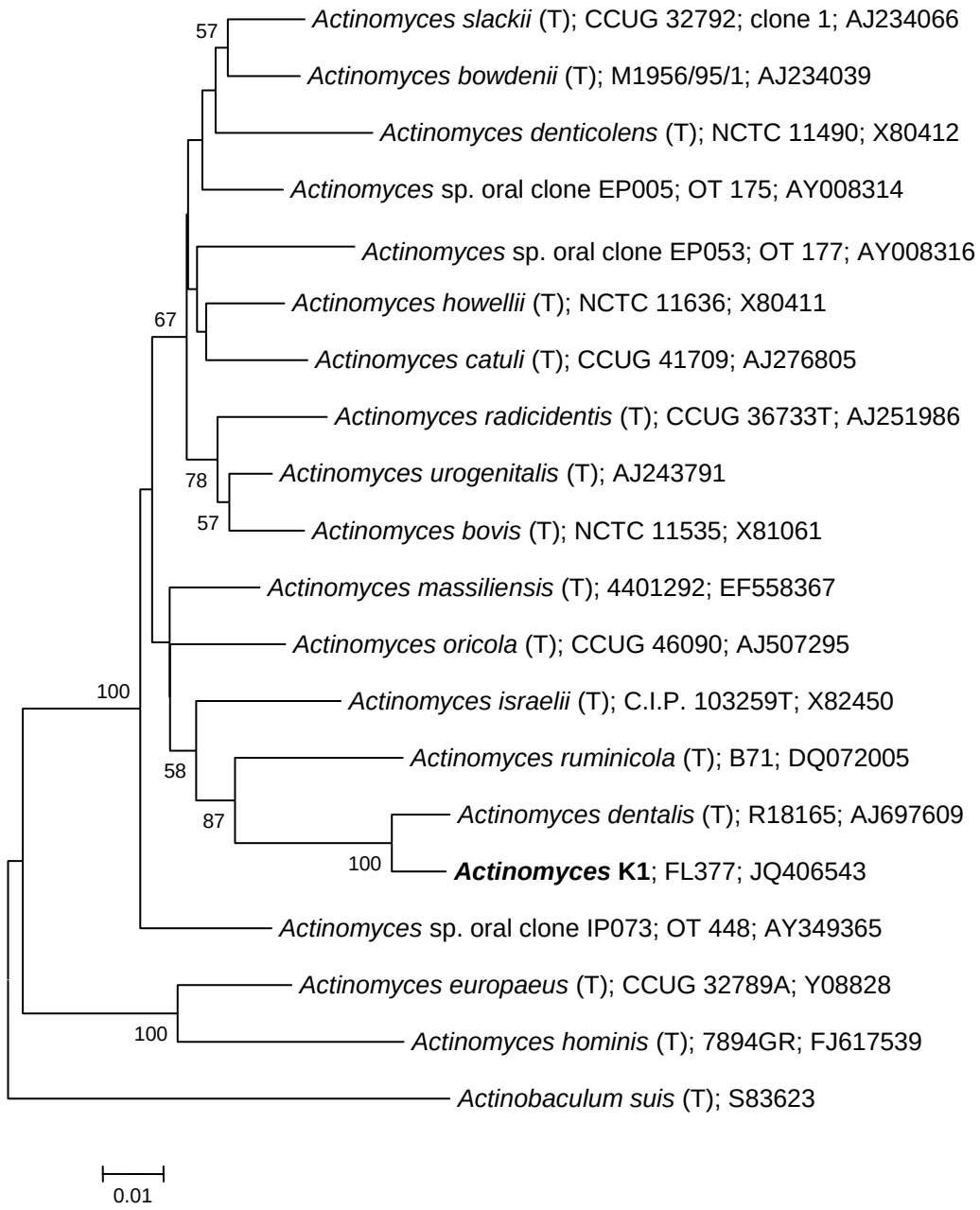


Figure 3.21: Phylogenetic tree based on 16S rRNA gene sequence comparisons over 1208 aligned bases from 20 nucleotide sequences showing the relationship between *Actinomyces* K1 and representatives from the genus *Actinomyces*. The tree was constructed using the neighbour-joining method from a distance matrix constructed from aligned sequences using the Jukes-Cantor correction. Numbers represent bootstrap values for each branch based on data for 1000 trees. Accession numbers for 16S rRNA gene sequences are given for each strain. Scale bar shows number of nucleotide substitutions per site.

3.4.3 Microbial community analysis of the Sanger-based sequencing data

All libraries were quality controlled by manually reviewing sequencing quality and annotating nucleotides where needed. All libraries contained the same number of sequences.

3.4.4 General observations and Good's coverage

Table 3.6 shows data for the libraries with data combined across patients for each library, detailing number of sequences, number of observed species and corresponding Good's coverage. Goods coverage describes how well the library represents the total community given varying criteria of uniqueness and can aid in the decision to extend sampling effort.

Number of observed species range from 86 for the isolates to 127 for the libraries. Generally, S_{obs} for the clone libraries were higher than the isolate values and ranged from 109 – 127. Goods coverage values for the individual libraries ranged from 81-86 %, while the value for the combined data resulted in a coverage rate of 98 %. Contrary to S_{obs} values coverage values for the clone libraries (81 – 86 %) were lower than that for the isolates (88 %).

Table 3.6: Libraries combined across patients detailing number of sequences (nseqs), number of observed species (S_{obs}) and Good's coverage (coverage)

Group	nseqs	S_{obs}	Coverage (%)
Isolates	540	86	88
Lib1	540	120	81
Lib2	540	127	81
Lib3	540	109	85
Lib4	540	121	82
Lib5	540	110	86
all	3240	229	98

3.4.5 EstimateS computed results

Table 3.7 shows the results of the EstimateS analysis, showing number of sequences, number of observed species (S_{obs}), Chao1, an estimate of total species richness, Inverse Simpson Index of diversity and ICE, a coverage-based richness estimator.

Chao1 values range from 137.04 for the isolates to 225.53 for the clone libraries. Mirroring S_{obs} the Chao1 values of the clone libraries are higher than the isolate values and range from 152.00 – 225.53. The Chao1 value for the combined data set is 275.41. In the case of the Inverse Simpson Index values range from 19.60 – 29.94. In this instance, the value for the isolates is very similar to the clone libraries, with the exception of library 5, which is much higher at 29.94 compared to the other libraries (19.60 – 22.59) and the value for the combined data set (25.37), but not significantly so. The coverage-based richness estimator ICE again mirrors the trend of the Chao1 data, isolates representing the lowest value (177.24), while the clone library values range higher (207.16 -319.58), as well as the value for the combined data (283.58).

Table 3.7: EstimateS data for patient data combined for each library showing nseqs, S_{obs}, ICE (incidence-based coverage estimator) and Chao1

Group	nseqs	S_{obs}	Chao1	InvSimpson	ICE
Isolates	540	86	137.04	19.60	177.24
Lib1	540	120	167.05	21.33	230.15
Lib2	540	127	192.33	19.97	319.58
Lib3	540	109	172.28	22.03	217.16
Lib4	540	121	225.53	22.59	263.93
Lib5	540	110	152.00	29.94	207.16
all	3240	229	275.41	25.37	283.58

3.4.6 Mothur computed results

Table 3.8 and Table 3.9 show a summary for mothur computed data for the combined patient data for each library and for all the individual libraries, respectively. The number of sequences (nseqs), number of observed species (S_{obs}), (Good's) coverage and values for the inverse Simpson Index of Diversity, which indicates the diversity while taking into account both species richness (number of species per sample) and species evenness (relative abundance of species making up the richness).

S_{obs} values range from 99 – 135, whereas the values for the individual libraries range from 5 – 51, showing the lowest average for patient A, while patient B has the highest number of S_{obs} on average. S_{obs} values for the combined data is 319 and 320 for the total of the individual libraries.

Coverage for the combined data ranges from 87 – 90 % for the libraries and 96 % for the total, while the individual libraries display a far wider range of 61 % (patient B Iso) – 99 % (Patient A Iso). The total of the individual libraries corresponds with that of the total for the combined data at 96 %.

Chao1 values for the combined patient data range from 169.50 for the isolates to 225.71 for the clone libraries. Mirroring S_{obs} the Chao1 values of the clone libraries are higher than the isolate values and range from 173.10 – 225.71. The Chao1 value for the combined data set is 506.64. Chao1 values for the individual libraries range from 5.00 (Patient A Iso) to 122.25 (Patient B L4). Patient A has the lowest average and patient B the highest. The Chao1 value of the total for the individual libraries is 516.57.

In the case of the Inverse Simpson Index for the combined patient data values range from 19.93 – 31.38. The value for the isolates is very close to those of the clone libraries, except for library 5, which is much higher at 31.38 compared to the other libraries (19.93 – 23.77) and the value for the combined data set (25.48). As for the other measurements, the Invsimpson values for the individual libraries show a far wider spread and range from 2.01 (Patient A L4) to 59.50 (Patient B L1). The total for the individual libraries is again similar to the total of the combined patient data at 25.88.

Table 3.8: mothur summary for combined patient data for each library

group	nseqs	S_{obs}	Coverage (%)	Chao1	invsimpson
Isolates	501	99	90	169.50	19.93
Lib1	501	130	88	204.39	20.98
Lib2	512	135	87	219.37	20.37
Lib3	511	122	89	173.10	23.77
Lib4	494	125	88	255.71	21.77
Lib5	524	128	88	234.11	31.38
all	3043	319	96	506.64	25.48

Table 3.9: mothur summary for individual libraries

Group	nseqs	Coverage (%)	S_{obs}	invsimpson	Chao1
A_isolates	90	99	5	2.41	5.00
A_Lib1	90	97	11	2.79	12.00
A_Lib2	87	97	9	2.34	10.00
A_Lib3	87	97	12	3.19	12.75
A_Lib4	86	95	10	2.01	13.00
A_Lib5	90	92	17	4.47	21.20
B_isolates	88	61	50	43.50	130.14
B_Lib1	85	65	51	59.50	78.19
B_Lib2	88	67	51	42.53	73.56
B_Lib3	84	70	42	30.58	69.27
B_Lib4	85	69	41	33.06	122.25
B_Lib5	81	81	35	23.31	43.08
C_Isolates	88	90	21	10.57	28.20
C_Lib1	90	79	38	22.13	52.25
C_Lib2	90	88	25	10.88	34.17
C_Lib3	89	92	23	14.61	26.50
C_Lib4	89	88	28	12.39	39.00
C_Lib5	90	91	22	10.11	29.00
D_Isolates	84	86	26	13.31	42.50
D_Lib1	89	70	47	38.77	72.07
D_Lib2	90	73	42	23.42	69.60
D_Lib3	90	79	37	10.71	50.15
D_Lib4	90	74	42	26.18	70.11
D_Lib5	90	71	40	20.43	94.17
E_Isolates	89	94	13	5.66	16.33
E_Lib1	87	91	14	3.03	28.00
E_Lib2	89	84	19	3.34	64.50
E_Lib3	87	89	20	4.93	29.00
E_Lib4	86	87	21	5.38	30.17
E_Lib5	90	83	26	8.47	52.25

Group	nseqs	Coverage (%)	S_{obs}	invsimpson	Chao1
F_Isolates	90	79	28	5.52	70.75
F_Lib1	89	80	30	11.16	60.60
F_Lib2	90	73	39	15.64	94.20
F_Lib3	90	80	33	12.10	48.30
F_Lib4	90	77	35	17.72	70.00
F_Lib5	90	77	35	11.64	65.00
All	3177	96	320	25.88	516.57

3.4.7 Libshuff Analysis

The Libshuff analysis uses the Cramer-von Mises test statistic. The significance of the test statistic indicates whether communities have the same structure. Due to the pairwise comparison the Bonferroni's correction was applied. Results calculated with mothur (Table 3.10) revealed only libraries 1 and 2 to be significantly different from each other. Libraries are considered significantly different if the significance value is ≤ 0.001667 according to p value of 0.05 divided by number of samples (in this case 30).

Table 3.10: Libshuff analysis comparing libraries for significant difference, significance highlighted with *

Comparison	dCXYScore	Significance
Isolates-Lib1	0.00201204	0.068
Lib1-Isolates	0.0011764	0.1921
Isolates-Lib2	0.00109894	0.1702
Lib2-Isolates	0.00103851	0.1865
Isolates-Lib3	0.00490316	0.005
Lib3-Isolates	0.0012341	0.2912
Isolates-Lib4	0.00166464	0.3802
Lib4-Isolates	0.00323136	0.1902
Isolates-Lib5	0.00805829	0.0019
Lib5-Isolates	0.00111779	0.5221
Lib1-Lib2	0.00204966	0.0491
Lib2-Lib1	0.00475098	0.0012*
Lib1-Lib3	0.00313374	0.0465

Comparison	dCXYScore	Significance
Lib3-Lib1	0.00289131	0.0735
Lib1-Lib4	0.00086047	0.7719
Lib4-Lib1	0.00587313	0.0807
Lib1-Lib5	0.00373116	0.1038
Lib5-Lib1	0.00362539	0.1782
Lib2-Lib3	0.00177326	0.1377
Lib3-Lib2	0.0021504	0.114
Lib2-Lib4	0.00091363	0.629
Lib4-Lib2	0.00630474	0.0335
Lib2-Lib5	0.00311455	0.0774
Lib5-Lib2	0.00231974	0.1839
Lib3-Lib4	0.00055972	0.6785
Lib4-Lib3	0.00329212	0.0759
Lib3-Lib5	0.00094408	0.3062
Lib5-Lib3	0.00193461	0.1046
Lib4-Lib5	0.00392276	0.015
Lib5-Lib4	0.00090808	0.4219

3.4.8 Rarefaction curves

Rarefaction analysis allows estimation of species richness as a function of number of samples, for this rarefaction curves are calculated by using a re-sampling without replacement approach. Using mothur rarefaction curves were computed for each library and patient, as well as the rarefaction curves for each library for the combined patient data and the rarefaction curve for all sequences combined, and are shown in Figure 3.22 to Figure 3.29. All of the curves show levelling off to a certain degree, with the curves of patient B and the curves representing library 1 of patient C and D, showing still the steepest slope. For patient B curves representing libraries Isolates, 1 and 2 are very close to each other, while curves for libraries 3 and 4 level off a little sooner, but show nearly the same trend, thus overlapping for about 2/3 of the curve. The same can be observed for patient E and curves representing libraries 3 and 4 as well as isolates and library 1.

The curves of libraries 2 and 4 of patient D overlap for the most part, making it difficult to see the actual curve for library 2. The same can be observed for libraries 4 and 5 for patient F.

The rarefaction curve for the combined patient data (Figure 3.28) shows the curve for isolates to be less steep than those for the clone libraries, which are all very similar. As could be observed in the individual patient curves, some curves overlap. In this case libraries 1 and 2 are congruent until nearly the end of the curve. The curves for libraries 3, 4 and 5 are less steep, but also overlapping. Libraries 4 and 5 are totally congruent, whereas the curve for library 3 levels off sooner than the other two.

The rarefaction curve of all data combined (Figure 3.29) shows a definite trend of levelling off and the lower and upper confidence intervals appear to be very close to the main curve indicating that the majority of taxa were discovered with the described sampling effort. Confidence intervals of the rarefaction curve display the range in which 95% of all re-sampled rarefaction curves will fall. Confidence interval curves can also be thought of as the upper and lower values of error bars displayed as a continuous curve rather than an individual error bar for each data point.

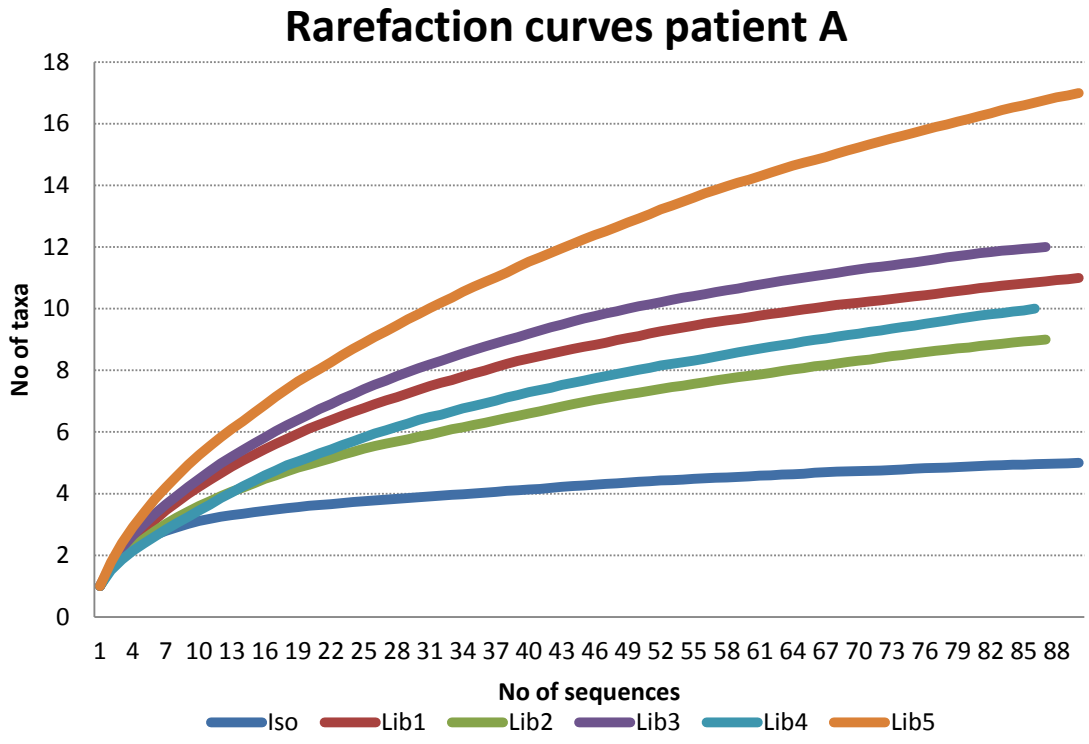


Figure 3.22: Rarefaction curves of microbial populations from dentine caries lesions in patient A using culture and five molecular libraries. The numbers of different OTUs (Operational Taxonomic Unit) in each sample are plotted vs. the number of sampled sequences.

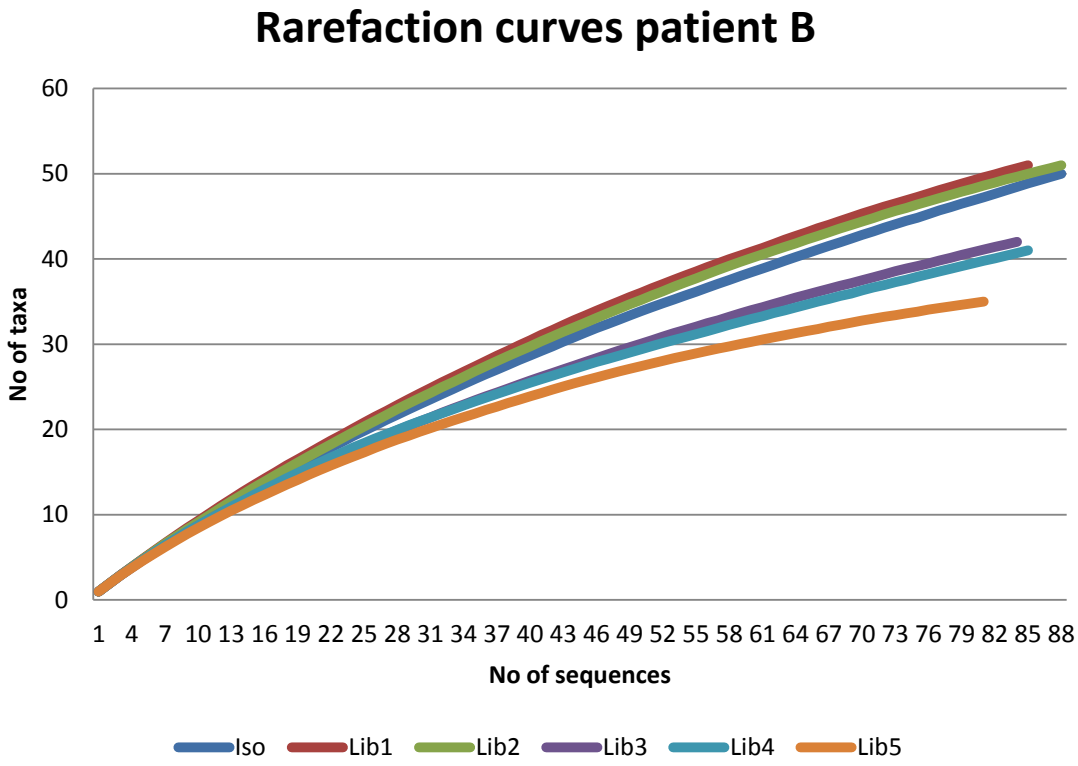


Figure 3.23: Rarefaction curves of microbial populations from dentine caries lesions in patient B using culture and five molecular libraries. The numbers of different OTUs (Operational Taxonomic Unit) in each sample are plotted vs. the number of sampled sequences.

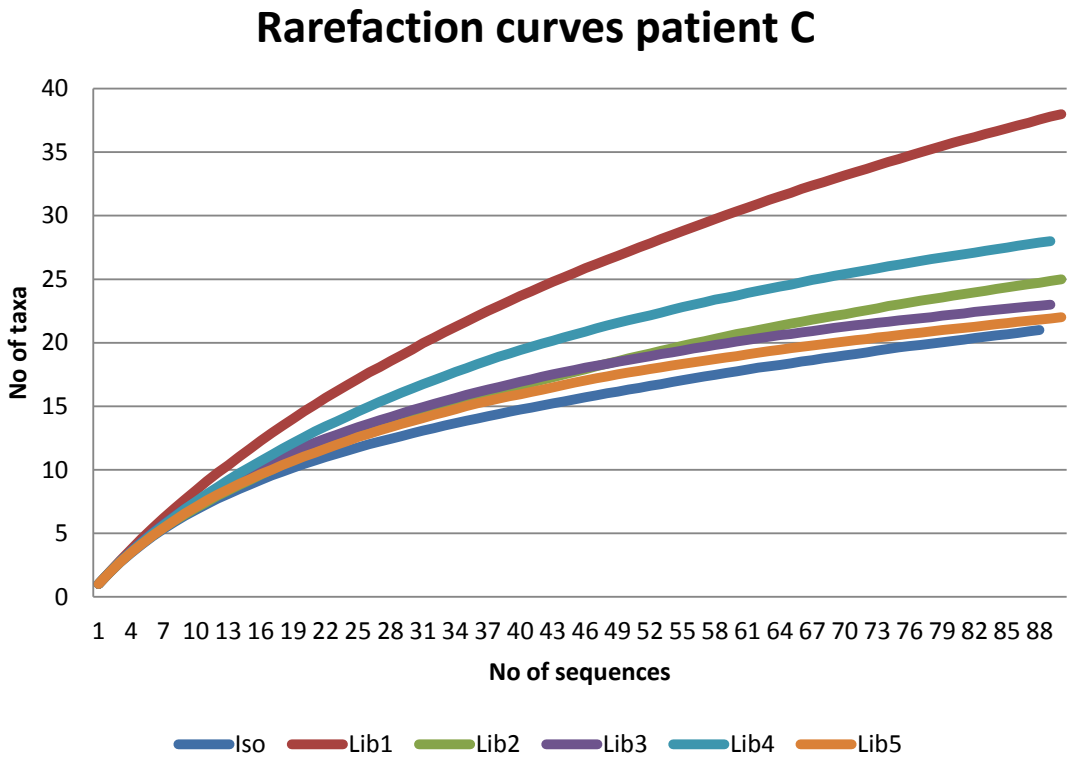


Figure 3.24: Rarefaction curves of microbial populations from dentine caries lesions in patient C using culture and five molecular libraries. The numbers of different OTUs (Operational Taxonomic Unit) in each sample are plotted vs. the number of sampled sequences.

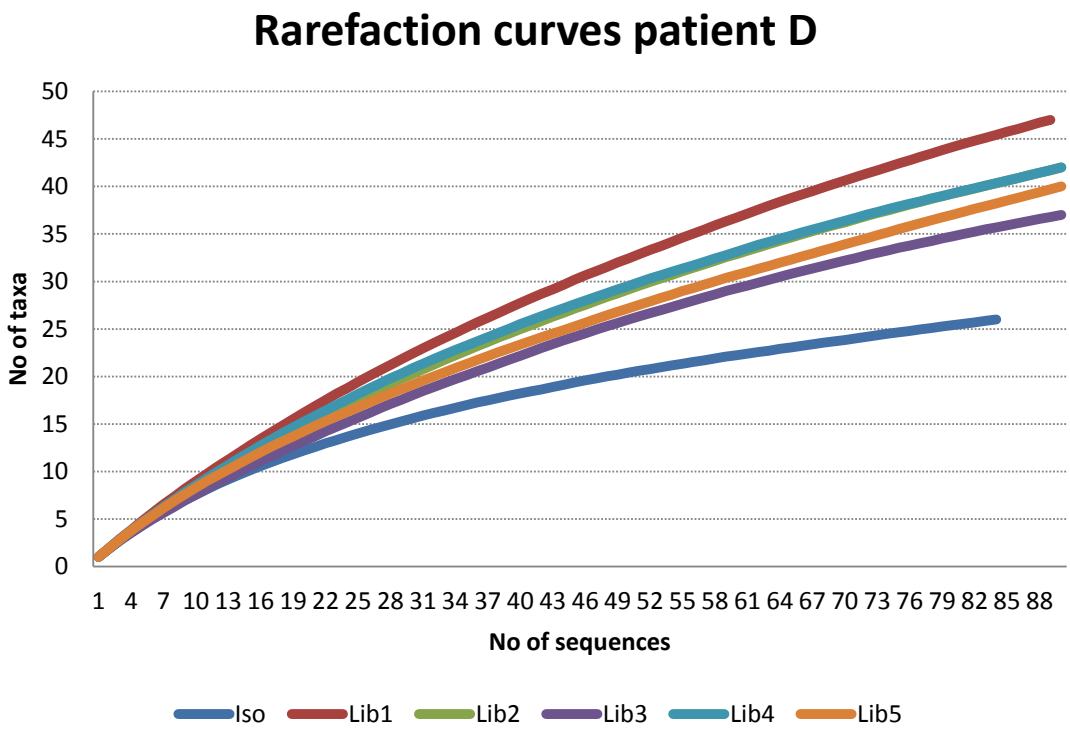


Figure 3.25: Rarefaction curves of microbial populations from dentine caries lesions in patient D using culture and five molecular libraries. The numbers of different OTUs (Operational Taxonomic Unit) in each sample are plotted vs. the number of sampled sequences.

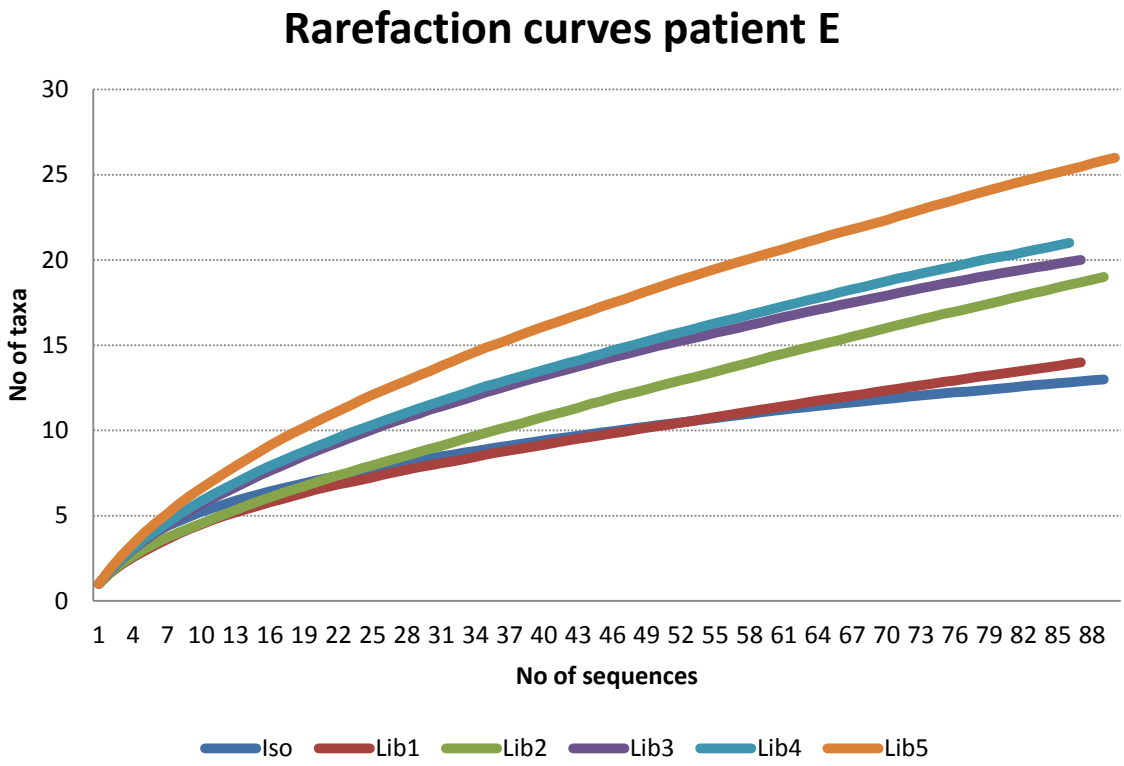


Figure 3.26: Rarefaction curves of microbial populations from dentine caries lesions in patient E using culture and five molecular libraries. The numbers of different OTUs (Operational Taxonomic Unit) in each sample are plotted vs. the number of sampled sequences.

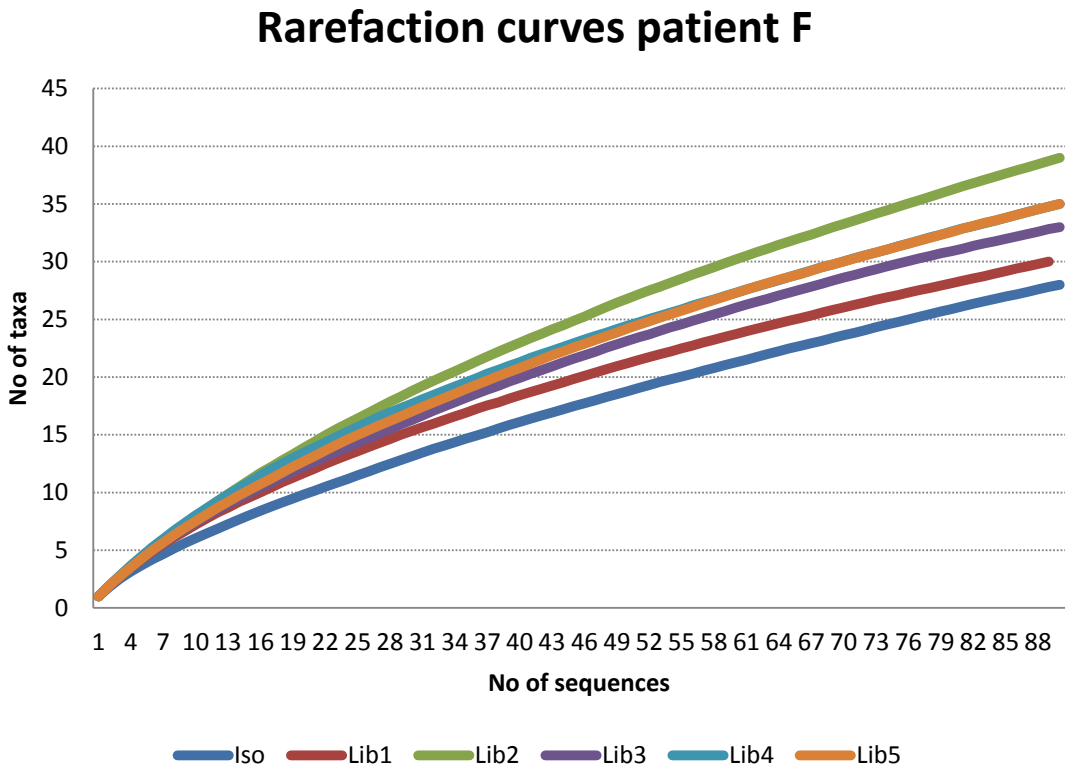


Figure 3.27: Rarefaction curves of microbial populations from dentine caries lesions in patient F using culture and five molecular libraries. The numbers of different OTUs (Operational Taxonomic Unit) in each sample are plotted vs. the number of sampled sequences.

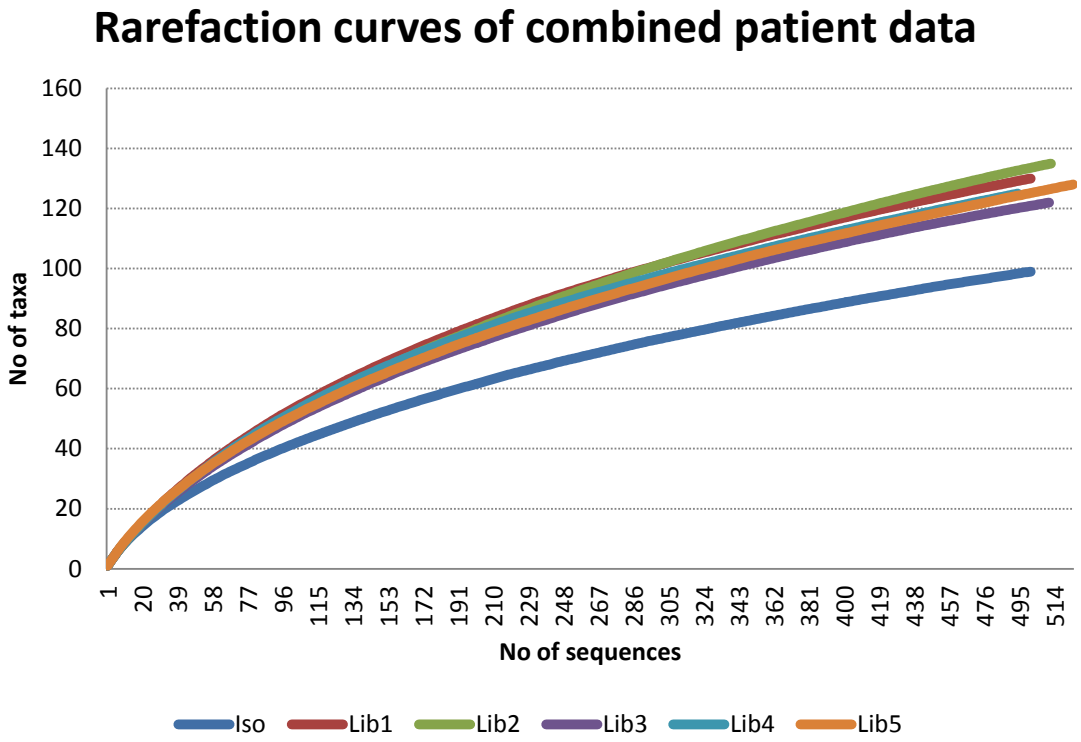


Figure 3.28: Rarefaction curves of microbial populations from dentine caries lesions in all patients combined using culture and five molecular libraries. The numbers of different OTUs (Operational Taxonomic Unit) in each sample are plotted vs. the number of sampled sequences.

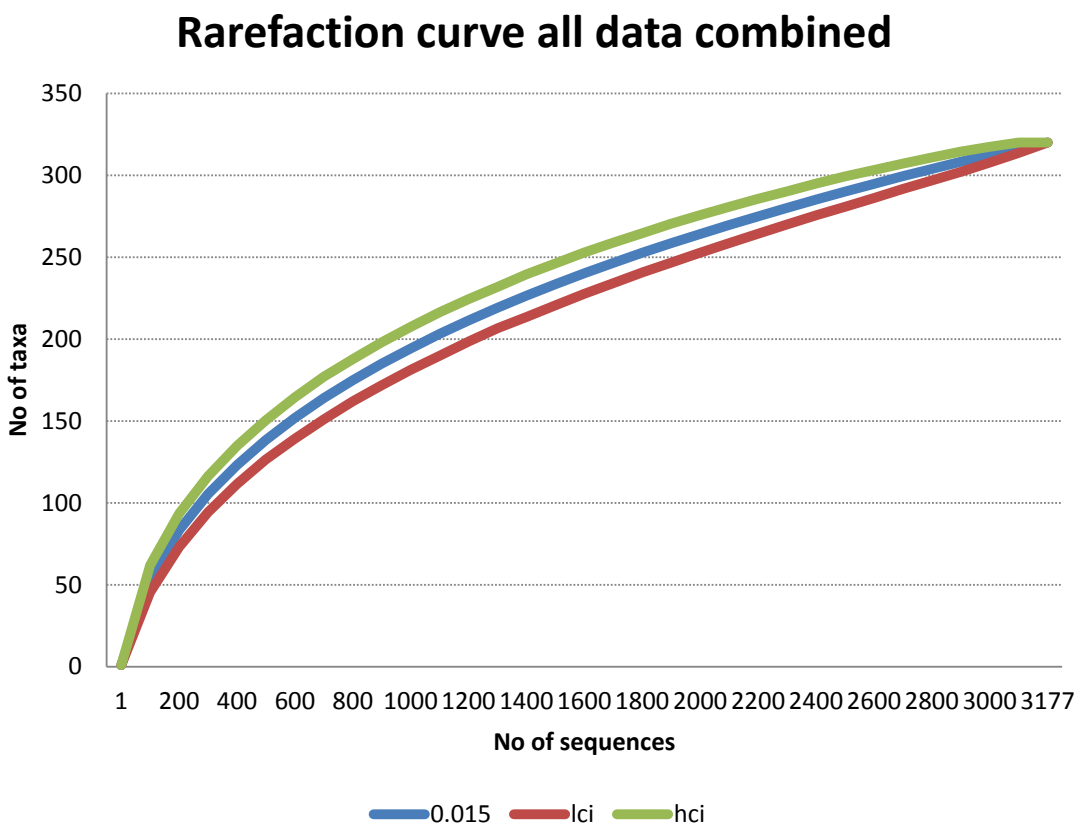


Figure 3.29: Rarefaction curve of microbial populations from dentine caries lesions of all patients and libraries combined using culture and five molecular libraries. The numbers of different OTUs (Operational Taxonomic Unit) in each sample are plotted vs. the number of sampled sequences. Upper (hci) and lower (lci) confidence intervals are also displayed.

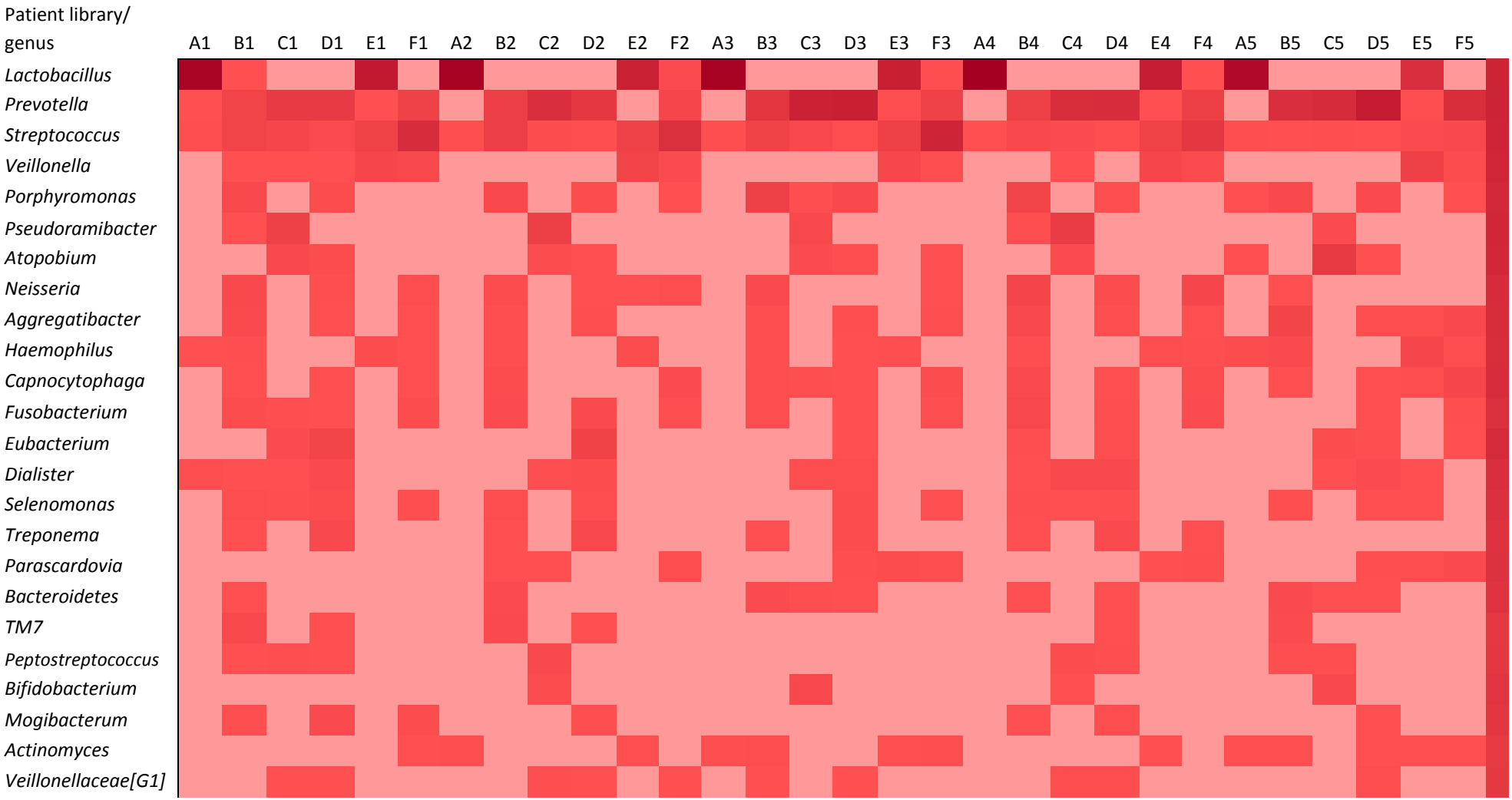
3.4.9 Heatmaps

Heatmaps in Table 3.11 and Table 3.12 show the relative abundance of the 50 highest-ranking genera of all the patient libraries. Table 3.11 depicts the numbers of the clone libraries, it can be seen that the top three genera make up a great proportion of the total. The highest-ranking genus shows great variation and illustrates how libraries from patients A and E are dominated by lactobacilli, whereas the other patient libraries show little or no incidence of this genus. The genera *Prevotella* and *Streptococcus*, ranking second and third place, show a more even distribution. The remainder of the table shows great variation among patients and libraries, albeit in relatively low numbers.

The heatmap for culture analysis shows slight differences to that of the molecular analysis. The highest-ranking genus is *Lactobacillus*, but the second and third ranking genera are *Olsenella* and *Prevotella.* Lactobacilli again make up a great proportion of the total, but as with the molecular analysis, a great variation in detection can be seen,

in that patient A and E have a high incidence, whereas no lactobacilli were detected in the other patient samples. In fact, among the 50 highest-ranking genera only two genera were detected from the isolates of patient A. All other patients showed greater diversity. Of the total 50 highest-ranking taxa, genera were detected for only 40 by culture.

Table 3.11: Heatmap of the 50 highest-ranking genera found in the 6 patient samples using molecular analysis



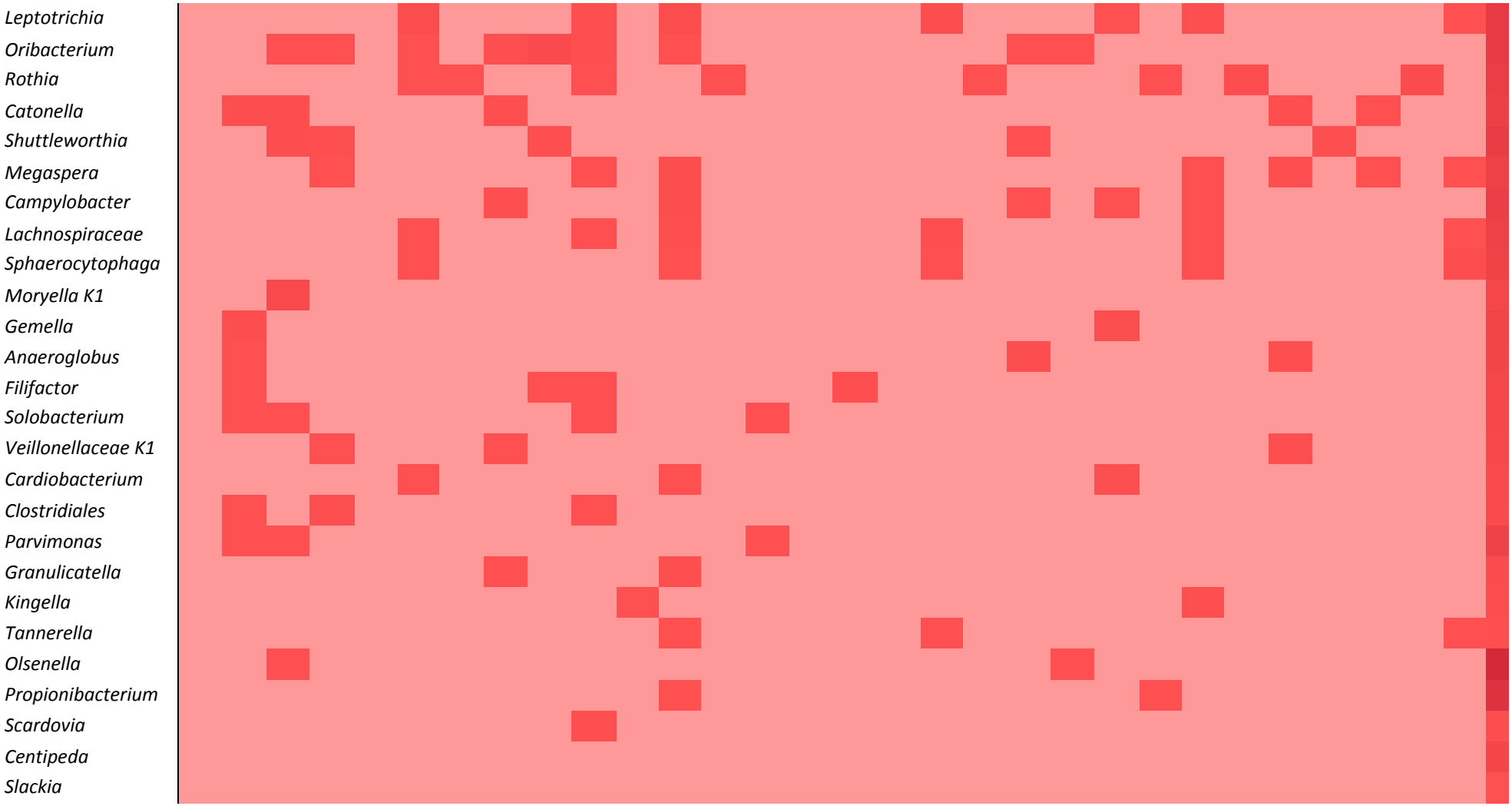
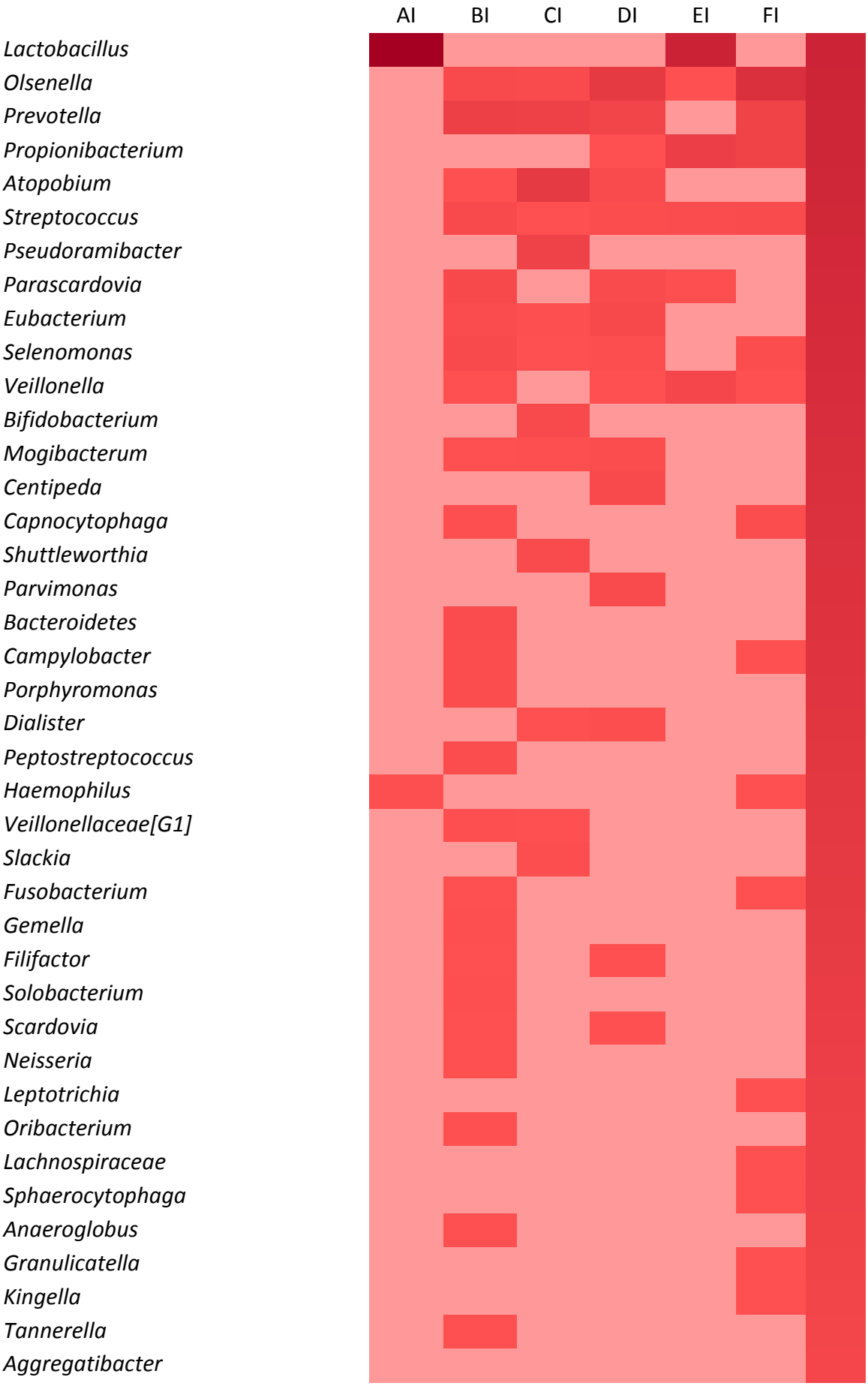

Leptotrichia
Oribacterium
Rothia
Catonella
Shuttleworthia
Megaspera
Campylobacter
Lachnospiraceae
Sphaerocytophaga
Moryella K1
Gemella
Anaeroglobus
Filifactor
Solobacterium
Veillonellaceae K1
Cardiobacterium
Clostridiales
Parvimonas
Granulicatella
Kingella
Tannerella
Olsenella
Propionibacterium
Scardovia
Centipeda
Slackia

Table 3.12: Heatmap of the 50 highest-ranking genera found in the six patient samples using culture analysis



Treponema
TM7
Actinomyces
Rothia
Catonella
Megasphera
Moryella K1
Veillonellaceae K1
Cardiobacterium
Clostridiales

3.4.10 Novel *Selenomonas*

It proved difficult to obtain good quality sequences for some isolates and clones, particularly isolates F061, F062 and F063. The sequences for these isolates appeared to be mixed with two or more peaks for the same position. To resolve this, 16S rRNA gene amplicons from single organisms were cloned and individual clones sequenced. Alignment of the resulting sequences showed marked inter-operon variation in the 16S rRNA sequence. Each strain exhibited three variants of the gene (Figure 3.30). Three clones from isolate F062, representing all three variants, were subsequently sequenced in full length. Alignment of all available sequences of good quality for F061, F062 and F063 with all other *Selenomonas* sequences revealed further isolates aligning according to the variants detected in F061 – F063.

Aligning sequences from F061 clones 1-16, F062 clones 1-16, F063 clones 1-16, F061 01 D, F062 01 contig, F062 08 D, F062 08 contig, F062 10 D, F062 10 contig, C085 clones 1-8, D033 clones 6-8, D035 clones 1-8 and B061 clones 1-8 showed all sequences to group into 6 variants as can be seen in Table 3.13.

Variant 1 is the longest at 563-575 bases long, with the exception of C085 clones 6 and 8, which showed a 93 base deletion at the beginning. Clones C086_06 and _08 also have individual base variations at the beginning and between bases 159 and 217 of the alignment.

Variant 2 is between 469-473 bases long and shows a relatively long deletion from bases 42 – 135 on the alignment, followed by a much shorter deletion at 143 – 146

bases. Regions 136 – 168 and 255 – 338 of the alignment show multiple base changes compared to variant 1.

Variant 3 470 – 474 bases long, shows the same deletion at the beginning as variant 2 and also shows the same base changes in region 136 -168. There is a region of base changes from 255 – 338 bases, but there is a great deal of variability within this variant. The only constant is a change from A to G at position 255 and T to C at position 315 compared to variant 1.

Variant 4 is 472 – 475 bases long and shows a slight difference at the beginning in that there is only one deletion from base 42 – 138 of the alignment. As seen with the previous variants base changes occurred in the regions 139 – 169 and 255 – 338, however, in this instance the base changes in the former region are different to those of the first two variants, whereas the changes in the first part of the second region (255 – 279) and at nucleotide position 338 are very similar to that of variant 2, while changes at 295 – 297 and 315 are similar to those of variant 3.

Variant 5 is 472 – 475 bases long and the least changed compared to Variant 1. The deletion at the beginning from bases 42 – 138 and changes in the region of 139 – 169 are the same as in variant 4. The changes at position 255 and 338 are also the same as for variants 2 – 4, but within this region only very few changes are observed.

Finally, variant 6 is 470 – 475 bases long and shows exactly the same alterations as variants 4 and 5, except that at position 198 of the alignment an A changed for a G, which was not observed with any other variant. Furthermore, there are base changes at positions 275, 277 and 322 that are unique to this variant, while changes at positions 297 and 315 are the same as in variants 4 and 3, respectively.

The various sequences were grouped together and are shown in Table 3.13.

```
F06201    5′ ..GAAATTAAAA [...] AGTCGAAGGGTTGAATGGTAATTTTTT---AATTTAGT [...] TGAAACGGAGGAGA 3′

F06208    5′ ..---------- [...] ----GAGC.AA-....A.C.TGC....AT-G.GC.... [...] ....GTCAGA.GAG 3′

F06210    5′ ..---------- [...] ----G..T.ATC.T..A.C.TGC....AGAG...C... [...] ....G......... 3′


5′ GGCATCTCTCTTCCGTGAAAGATGGCCTCTATTTATAAGCTATCACCTGTTGATGGGTCTGCGTCTGATTAGCTAGTTGGTGAGGT... 3′

5′ ............................T.....--.....................................G.............. 3′

5′ ..............................................................................G.............. 3′
```

Figure 3.30: Alignment of the three *Selenomonas* variants found in isolate F062. Contig regions of the 16S rRNA gene of clones 01, 08 and 10 of isolate F062 sequenced in near full length showing sequence variation are highlighted.

Table 3.13: *Selenomonas* sequence variants listed by their respective clone sequences

Variant	Isolate number	Clone number
1	F061	02, 07, 09-12, 15, 16
	F062	01, 06, 07, 12, 14
	F062	01 D and 01 contig
	F063	01, 02, 04, 06, 10, 12-16
	C085	06, 08
2	F061	14
	F062	03, 08, 11, 13
	F062	08 D and 08 contig
	F063	05, 07, 09, 11
3	C085	01-05, 07
	B061	02-08
4	D026	03, 05, 06, 08
	D035	01, 02, 04-06
5	B06	01
	F061	01, 08, 13
	F062	04, 09, 10, 15, 16
	F062	10 D and 10 contig
	F063	03, 08
6	D026	01, 02, 04, 07
	D033	04, 06-08
	D035	03, 07, 08

Figure 3.31 shows a phylogenetic tree using full length sequence contigs of Selenomonas sequences detected in patient samples. The contig variants found in patient F are highlighted in bold.

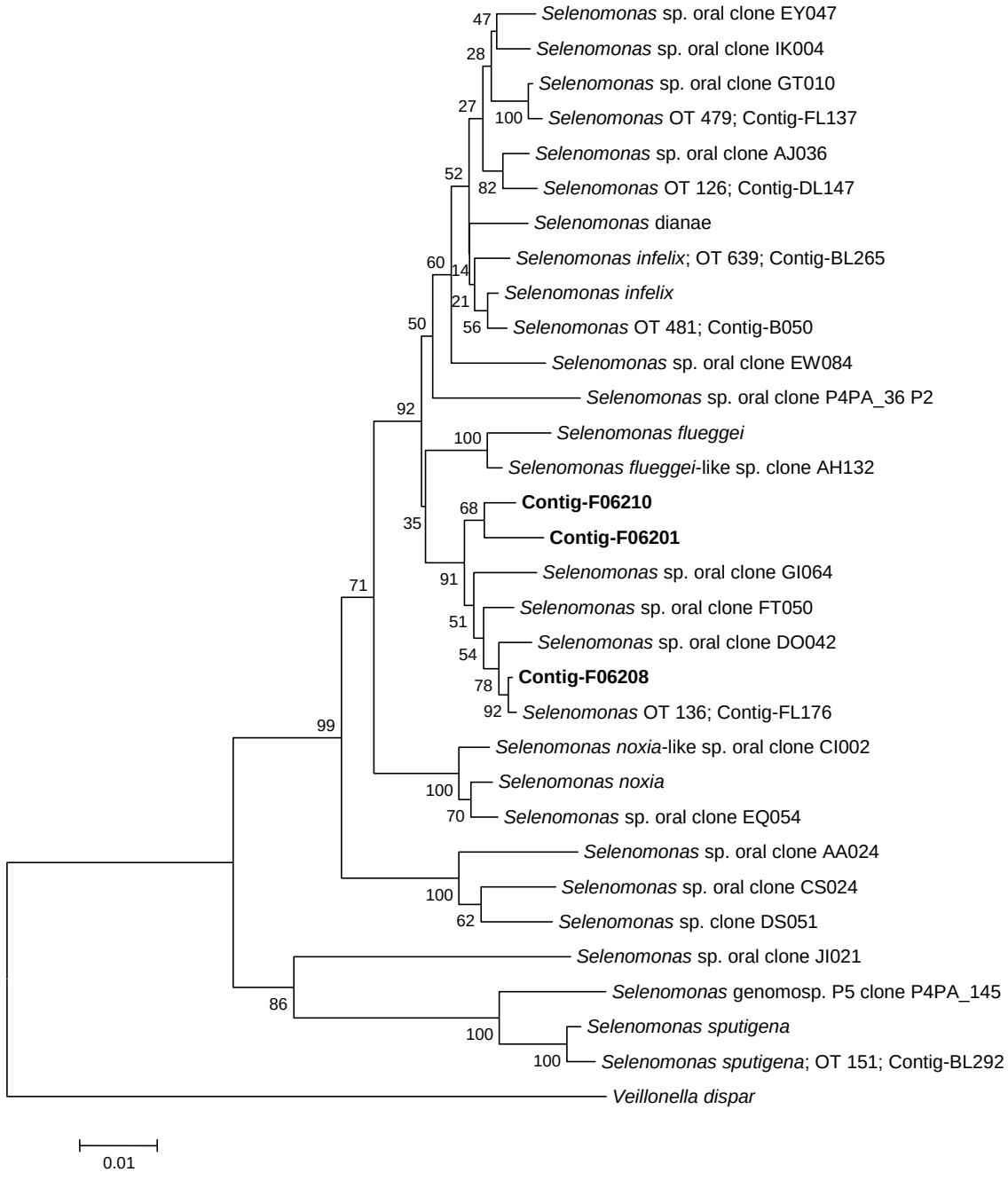


Figure 3.31: Phylogenetic tree based on 16S rRNA gene sequence comparisons over 1290 aligned bases from 32 nucleotide sequences showing the relationship between the three *Selenomonas* variants initially detected, other full length *Selenomonas* contigs from patient samples and representatives from the genus *Selenomonas*. The tree was constructed using the neighbour-joining method from a distance matrix constructed from aligned sequences using the Jukes-Cantor correction. Numbers represent bootstrap values for each branch based on data for 1000 trees. Accession numbers for 16S rRNA gene sequences are given for each strain. Scale bar shows number of nucleotide substitutions per site.

3.5 Discussion

In this chapter, six patient samples, from soft dentine caries lesions, were collected and the microbial load analysed using culture and molecular techniques. Five primer sets, some of them novel and designed as part of this study, were used in the molecular analysis, and isolates were grown from each sample. A wide diversity of bacteria were detected in inter- and intra-library and –patient comparisons. Novel sequences were detected with every primer pair and sequences that were difficult to sequence revealed recombination amongst *Selenomonas* species not previously observed. Different analytical methods were applied, and all were in agreement that good sampling coverage was achieved.

One of the aims of this work was to improve the detection of high G+C species in the molecular analysis and the work described in Chapter 2 suggested that the novel primers (27F YM, 39F and 61F) had potential to achieve this. The results of this chapter showed that none of the five primer pairs used increased detection of high G+C species to the detection level of the cultural analysis.

Examining the counts for the culture analysis of the six patient samples revealed patient A to have the lowest number of colony forming units at 5.4×10^4, whereas counts for the other five patient samples ranged between 1.2×10^7 and 2.0×10^8. This demonstrates possible bias, in that the samples were a scoop taken from the caries lesion, which of course is individual in nature, i.e. caries lesions were at slightly different stages of progression through the dentine and the volume and weight of each scoop was not recorded or normalised. This was done in the study by Chhour et al. (2005), but omitted from the protocol in this study to ensure samples were subjected to as little oxygen as possible and thereby increasing the chance of detecting fastidious anaerobes that might otherwise die.

The target of improving the detection rate of the high G+C phylum Actinobacteria was not reached and the results were similar to those from previous studies (Munson et al. 2002; Munson et al. 2004; de Lillo et al. 2006), in that the proportions of *Actinobacteria* in the molecular analyses were reduced compared to culture. In Chapter 2 analysing a model community, it was shown that primers 39F and 61F gave improved recovery of the phylum *Actinobacteria* compared to different versions of the

commonly used primer 27F. Analysis of the patient samples reveals that molecular analysis still does not come close to the detection rate when culture analysis is used, although primer pair 61F / 1387R was again superior (11.1 %) compared with the other primer pairs (2.6 % - 6.1 %). Overall, 32.2 % of isolates were identified as *Actinobacteria* (174 out of 540), but only 5.3 % of clones (143 out of 2700) were identified as members of this phylum.

Another reason for the decreased detection rates of *Actinobacteria* using molecular analysis could be that detection was not actually dramatically decreased, but that proportions appeared lower since molecular methods detect unculturable species, thereby lowering the proportion of *Actinobacteria* / detection in comparison to culture, which detected fewer taxa (90 vs. 216). Admittedly this explains the under-detection probably only in part, but should be borne in mind when comparing culture and molecular analysis.

All 57 *O. profusa* and 19 of the 21 *O. uli* were detected using culture. This was surprising since *O. uli* was included in the panel of bacteria for the validation of PCR primers (Chapter 2) and although amplification product compared to that of the other species in the test panel was often less, good product was always obtained. This shows that even though amplification of pure, single species template may be achieved, this does not necessarily hold true for that same species in an environmental sample.

The different primers in their various combinations used for the molecular analysis of the carious samples exert different biases on the detection of various species. Although primers targeting 16S rRNA genes are generally referred to as 'universal' primers, it is now accepted that primers are not, in fact, universal, but should rather be called broad range primers, since regions of the 16S rRNA gene are highly conserved, rather than fully (Baker et al. 2003; Wang et al. 2009). In addition, new primers are designed using alignments of species known to date, which often contain sequencing errors (Ashelford et al. 2005) and consequently cannot include species with less conserved 16S rRNA gene sequences that are as yet undiscovered. This explains why the clone libraries contain different species, but only few novel species.

As demonstrated in Figure 3.2 the detection rate of streptococci was dependent on primer combination and primers 61F / 1387R showed a clear bias, detecting between

3.28 and 5.17 times less streptococci than any other primer combination. Even when using culture analysis 1.28 times more streptococci could be detected. Primer 61F was adapted from primer 63F to avoid a mismatch at the 3' end (as described in Chapter 2), but it seems the bias was not removed completely. It appears that factors intrinsic to the DNA sequence, but outside the target region are affecting amplification adversely, as has been discussed in Chapter 2.

Influence of the primers in representing the environmental microbiome was shown in two separate studies looking at the quantitative effect of a single primer-template mismatch using the 16S rRNA gene as the template. It was found, that single mismatches in the first 3 - 4 nucleotides from the 3' end can greatly reduce and even hinder extension (Bru et al. 2008; Wu et al. 2009). This effect was reduced the closer the single mismatch was positioned to the 5' end of the primer and it did not appear to be influenced by the genetic context, unlike the above discussed example.

Further evidence highlighting the primer bias and more specifically, the bias the forward or the reverse primer can impose on the amplification reaction is illustrated in Table 3.5. It can be seen that some species (*P. alactolyticus* and all detected belonging to the phylum *Fusobacteria*) showed a lower detection rate when reverse primer 1387R was used (libraries 3 and 5), compared to detection with reverse primer 1492R (libraries 1, 2 and 4). In contrast to this is the finding that when the sequencing data is examined at phylum level, it becomes apparent that reverse primer 1387R, which is used in libraries 3 and 5, enabled a higher rate of detection of *Bacteroidetes* than analysis using primer 1492R. In the case of *P. oralis* and *P. tannerae,* detection using 39F / 1387R is significantly increased compared to any other primer combination, but the effect a different forward or reverse primer can have, becomes clear when looking at results using 39F / 1492R and 61F/1387R, which had a 4.96 times lower detection rate in both cases for the former and 2.16 and 1.78 times lower detection rate, respectively, for the latter.

Detection of *Atopobium* OT416, *A. segnis* and *H. parainfluenzae* was also improved using primer combination 61F / 1387R. In this instance, the effect is most likely due to the influence of the forward primer, since for these species rates of detection with primer set 39F / 1387R were not as different compared with the clone libraries using

reverse primer 1492R than it was for 61F / 1387R. To be able to make such a comparison a library using the primer combination 61F / 1492R would be needed, but the application of this primer set to a panel of individual bacteria *in vitro* (Chapter 2, section 2.4.6) indicated decreased product yield compared to primer pair 61F / 1387R applied to the same species.

The results reported by Bru et al. (2008), who reported differences in the detection rates using forward and reverse primers with a single mismatch, could explain the observed differences in detection for the various primers. They found forward primers with a mismatch located more than four bases away from the 3' end underestimated gene copy numbers, while no effect was seen on the reverse primers. They conceded that the severity of this bias is determined by numerous factors, such as the primer length, the nature and position of the mismatches and the annealing temperature of the primers. Since it is hoped to further the detection of as-yet undetected species using the primers described here it is of course impossible to exclude the possibility that detection of novel species is prevented by mismatches in the crucial 3' region. However, it appears that neither the processivity of the polymerase nor the choice of primers can explain the bias against organisms with a high G+C content. The last possible factor for introducing this bias against the high G+C organisms seems to be the cloning reaction.

In Chapter 2 it was discussed how DNA extraction, DNA polymerases, choice of primers and additional factors such as hardware elements can cause a biased amplification. In this chapter, a further potentially significant bias was introduced by cloning the amplified patient samples. Palatinszky et al. (2011) looking at preferential ligation during TA-cloning of multi-template PCR products found that libraries from an environmental sample constructed using the cloning methods compared to a length heterogeneity PCR, not using any cloning, showed a significant difference. However, Taylor et al. (2007) found no biases due to TA cloning when testing phylogenetic bias in fungal environmental clone library construction comparing TA and blunt-ended cloning. The abundance of OTUs was found to be correlated and phylogenetic tests showed no significant differences between the two libraries (Taylor et al. 2007). What they did find was that some OTUs abundant in the one library were not in the other and *vice versa*; indicating a potential phylogenetic bias during cloning. Palatinszky et al.

(2011) offered a similar explanation when OTU ratios of mock communities were slightly lower than expected, saying that this could be caused by sequence differences, since amplicons were of a similar length and G+C content.

Palatinszky et al. (2011) argue that the different findings of the two studies could be due to the different characteristics of the insert pools, more specifically, the insert sizes used. Their insert sizes ranged between 430-550 bp, which is much shorter than the 1100-2000 bp used by Taylor et al. (2007) or the ~1500 bp sequence lengths used in this study. However, ratios of largest to smallest insert size are actually greater in the study by Taylor et al. (2007) and much smaller in this study, so it seems an unlikely explanation.

Palatinszky et al. (2011) furthermore argue that clone libraries with low diversity may be more prone to phylogenetic biases and that drawing conclusions from diverse communities such as used by Taylor et al. (2007) or in this study could lead to underestimation of the extent of the bias. This does not appear to be logical, since in libraries with high diversity this bias would not weigh as heavily and would appear reduced. It would of course have to be borne in mind when analysing data with regards to methodological bias.

Preferential ligation can certainly depend on other factors, such as length of incubation coupled with temperature during ligation, as well as sequence properties of the inserts and differences of the cloning systems (Wang et al. 2009). However, as long as the same cloning system and ligation conditions are used for every library for comparative analysis, the same (likelihood of) bias applies and although some sequences may be under- or over-represented, libraries can be compared amongst each other.

When comparing the distribution of taxa of all libraries combined for each patient it became apparent that when a patient has a high incidence of lactobacilli (patients A and E), few *Prevotella* species (A=1, E=8 sequences identified as *Prevotella*) and no *Pseudoramibacter* were observed and vice versa (patient F=8 sequences identified as lactobacilli). Furthermore, where the microbiota of two samples was dominated by lactobacilli the overall number of taxa was lower compared to the other samples (22 and 41 taxa versus 113, 60, 117 and 88 taxa).

Chhour et al. (2005) reported similar findings when looking at the microbial diversity in advanced caries lesions of 10 patients. They, too, reported *Lactobacillaceae* and *Prevotellaceae* to make up the majority of all identified sequences and that lesions could be grouped into *Prevotella*-dominated or *Lactobacillus*-dominated samples. They argued that colonisation, or exclusion thereof, could depend on fermentation by-products, but that lactate did not appear to be the major fermentation by-product, since they observed lactate dependent *Veillonella* spp. infrequently and no further species capable of metabolizing lactate. Again, the results of this study are congruent, especially in that lactate-utilising phylotypes, such as *Selenomonas* spp. and *Pseudoramibacter alactolyticus*, were detected in lesions high in *Prevotella* (Chhour et al. 2005). The conclusion from these observations are that the incidence of the dominant species depends on factors early in the colonisation of the dentine matrix and successive inclusion or exclusion of subsequent colonisers is likely to be determined by metabolic by-products of the initial colonisers (Byun et al. 2004; Nadkarni et al. 2004; Chhour et al. 2005).

Spread over the five primer sets used, 16 novel sequences representing novel species, genera and possibly a new order were detected in the four predominant phyla. It was not unexpected to find novel species, since it is estimated that only just over 60 % of oral species have been discovered (Parahitiyawa et al. 2010). It was surprising, however, that library 1, utilising the commonly used primer pair 27F CM / 1492R detected the most clones (10) that represented novel sequences. Library 3, using the newly designed forward primer 39F and reverse primer 1387R, came second with six clones representing novel sequences, while Library 2, using primer pair 27F YM / 1492R revealed five novel sequences. Thus, primer combination 27F (CM and YM) / 1492R was responsible for the discovery of 16 out of 29 clones representing the novel sequences. Reverse primer 1492R seemed to have a detrimental effect concerning amplification of as-yet undiscovered sequences, since library 4, using primers 39F and 1492R only detected 2 novel sequences, both of which were also detected using libraries 1 and 2. It might be interesting to create a new clone library using primers 27F (CM and/or YM) / 1387R to see if further novel sequences can be detected from the same patient samples.

Another approach to advance the detection of novel species could be to use the universal primers to detect new groups of bacteria, as was achieved in this study, and to subsequently design family/genus-specific primers to broaden the knowledge of that particular group. It was shown in several other studies that greater diversity could be observed using genus specific primers (Vartoukian et al. 2009; Lin et al. 2011; Xie et al. 2011)

Richness and evenness measures to characterise microbial community structure have become more and more commonplace in the field of microbial ecology. Since the diversity of the sampled habitat is unknown prior to sampling, but at the same time an appropriate sample size needed, which is dependent on the habitats' diversity, statistical analyses estimating species richness and diversity are employed (Shaw et al. 2008; Barriuso et al. 2011).

The S_{obs} (species observed) are the number of observed OTUs for an OTU definition, which in this study was set to 98.5%. The numbers for the general observations and EstimateS are the same, since the same results from BLAST searches on HOMD are the basis for calculations. The S_{obs} calculator on mothur consistently observed more species than EstimateS. Numbers varied between +3.30% (Lib 4) and +16.36% (Lib 5) for the culture and clone libraries and was 39.3% higher for the combined data. This is probably due to manually correcting sequencing error, which is not done for mothur, which therefore excludes sequences from the analysis using the screen.seqs command.

The S_{obs} in the mothur analysis for the individual libraries ranged from 5 (A Iso) to 51 (B Lib2) with an average of 28.83 and a standard deviation of 12.96. A clustering can be observed for five of the six libraries from patient A, which have the lowest S_{obs}, followed by 5 of the 6 patient E libraries and the remaining patient A library. The clustering then becomes less distinct, but a trend can be observed with increasing S_{obs} from patient C to F to D with patient B libraries showing the highest rate in S_{obs}. This could be an illustration on how every person has their distinct set of bacteria (Costello et al. 2009).

On the other hand, Libshuff analysis indicates the libraries not to be significantly different. Libshuff analysis run on mothur shows that libraries 1 and 2 show significant

statistical difference, however, this seems unlikely since primer sequences of the forward primer 27F CM and 27F YM differ only in one position, while being very different to 39F and 61F. The most likely explanation for this observation is that the application of the Bonferroni correction that controls for experiment-wise error and allows multiple comparisons of libraries is not infallible, especially since the value of 0.0012 indicating statistical significance is relatively close to the cut-off value of 0.00167 calculated with the Bonferroni correction.

The value for Good's coverage represents the estimated proportion of the population represented by the sequences obtained from the samples. The Good's coverage was calculated using the same formula for both, but mothur had higher values for S_{obs}, which is the basis for the calculation, and accordingly, values were higher by 2.27 – 8.64 %. The only value being smaller for the mothur analysis was the one for all samples combined (-2.04 %). This can be explained if the number of overall singletons is considerably higher than for the Good's coverage calculated using the general observations. Coverage values were between 81 % and 90 % for both analysis approaches, indicating a good sampling effort for all libraries. Values for the individual libraries had coverage rates as low as 61 % while others were as high as 99 %. From these results, re-sampling for some of the libraries would be recommended to ensure a more even distribution of coverage rates and thereby lowering the bias this could cause.

The Chao1 results represent the richness estimate for the OTU definition of 98.5 %. Chao1 is a measure of minimum richness in a community. The EstimateS results for Chao1 are all higher than S_{obs} (38.18 - 86.39 % for isolate and clone libraries, 20.27 % for all values combined). This, as well as the values for Good's coverage, suggest that more isolates and clones need to be analysed to capture the true richness of this habitat, especially considering that Chao1 is an estimator of the minimum richness. Chao1 values for the mothur analysis are higher than S_{obs} (41.80 – 104 % for libraries and 58.5 % for all data combined), but also higher by 3.44 - 56.9 % for libraries and 83.7 % for the complete dataset than the Chao1 values for the general observations.

While richness describes the number of species in a sample, evenness describes the relative abundance of a species (making up the richness). Both terms together describe

the diversity of a sample and the Inverse Simpson Index of diversity describes the number of species present in a sample if they were evenly abundant. Values for this study for EstimateS were ranging between 19.6 – 29.94 for the libraries and 25.37 for the combined data set, while values for the mothur analysis were ranging from 19 – 31 for the libraries and 25 for the combined dataset. This shows that the richness of the patient samples in this study is quite high, but that evenness is rather low and therefore the diversity is low. This finding is in contrast to that of Shaw et al. (2008) who assessed the relative diversity of 16S rRNA libraries from the Global Ocean Survey and found that the diversity indices investigated yielded rankings very similar to the richness estimators. They concluded that sites relatively rich in taxa appear to also have relatively even taxon-abundance distributions. Although figures for S_{obs}, coverage and Chao1 values vary between the two analysis methods, results for Inverse Simpson diversity index were near identical for EstimateS and mothur.

ICE, the incidence-based coverage estimator, is a species richness estimator giving values for total species richness, including species not present in any sample. Values for each library and the combined data set ranged from 177.24 to 319.58 and were higher than the Chao1 values in every case. This is expected, considering that Chao1 represents minimum richness estimates while the ICE estimate covers species not present in the samples.

Rarefaction curves display the expected number of higher taxonomic groups, represented in a random selection of lower taxonomic units, such as species or number of sequences (Gart 1982). In this study, phylogenetic taxa were plotted against individuals. The shape of the curve illustrates a measure of evenness of diversity. For example, rarefaction curves for patient A show the curve for library 5 is steeper than that for the isolates, implying that for a given number of sequences, there are more taxa in library 5 than in the isolates library. However, while library 5 has more taxa per sequence, in the isolates library sequences are more evenly distributed within the taxa. This statement is supported by the S_{obs} value, which is lowest for isolates and highest for library 5 and also the Inverse Simpson diversity index, which is highest for library 5 and third lowest for isolates. All curves for patient A are starting to tail off, indicating a good sampling effort, which is supported by the coverage values of 92 – 99 % and the Chao1 values, which are very close to S_{obs} values (Table 3.9). Except for

patient B, isolates show high coverage rates and greatest evenness. Many of the curves for the various patients and libraries show no sign of tailing off, reflected by the lower coverage rates compared to those that do show a trend to tail off.

A comparison between the individual patient curves and the rarefaction curves for each library, for which the patient data was combined, show a good sampling effort for the libraries with coverage rates between 87 – 90 %. However, when considering data for the individual patients additional sampling for most of the libraries for many patients is indicated. Thus, when looking at the diversity and coverage rates at low resolution (i.e. all data combined Figure 3.29) a good sampling effort is seen, but when looking at the data in higher resolution (i.e. individual rarefaction curves for each patient) results could be improved by additional sampling.

Several points of caution have to be raised with regards to rarefaction curves. Rarefaction curves demonstrate the number of taxa in a sample and do not incorporate which taxa have been detected. Rarefaction curves from two samples of the same habitat that show a similar amount of taxa could have different species compositions and there is a risk of underestimating the habitat's richness if only the curves are evaluated. Therefore, rarefaction curves of the combined data should be compiled, or the taxon composition of the samples should be compared. And although the shape of the curve illustrates a measure of evenness of diversity, it does not account for the relative abundance of each species, like a true diversity index, such as Inverse Simpson.

The detection of the novel *Selenomonas* sequences and the observed recombination is very interesting. The concerned isolates were very difficult to sequence and appeared either as bad sequence reads throughout or started as a good quality read and then appeared to be mixed sequence. To resolve the issue the PCR product from the initial amplification was cloned and it was found that the different versions of the 16S rRNA gene within a single strain represented three different oral taxa in the HOMD database. This has implications for other studies, since it is likely that similar sequences have been disregarded as failed sequence reads and not investigated further, thereby missing this phenomenon. The initial three isolates that were difficult to sequence originated from the same patient and were picked subsequently from the initial

culture plate, as indicated by ascending isolate numbers. One could argue that this phenomenon is a one-off, but since five isolates from three further patients showed the same kind of recombination and resulted in six variants overall, it appears this phenomenon is more widespread. It is true that inter-operon variation is well known in bacterial species, however, the degree of sequence variation observed in this study is exceptional.

It is generally assumed that recombination / horizontal gene transfer that is tolerated and maintained provides a selective advantage. Several questions raised in this case, if all three variants in one host are functional and if one or all confer an advantage on the host. Could they allow some form of adaptability? Could they increase level of fitness in this particular habitat? Could it convey disease progression? Or could it even equip the host with some form of resistance, to ribosome-targeting antibiotics, for example?

These findings certainly highlight the need for caution when using the 16S rRNA gene for the identification of bacterial species. Hanage et al. (2005) addressed the ongoing debate on the concept of whether a universal species is possible for bacteria in a study looking at fuzzy species among recombinogenic bacteria. It is established that it is nearly impossible to achieve resolution down to species level with closely related, highly recombinogenic bacteria that colonise the same body site, such as *Neisseria, Actinomyces* or *Streptococcus* (Hanage et al. 2005; Do et al. 2009; Henssge et al. 2011). Multi locus sequencing analysis was used by Henssge et al. (2011) to analyse intra- and inter-species diversities of three *Actinomyces* species. Additional possible methods to achieve resolution that should be considered when trying to achieve higher taxonomic resolution when 16S does not give clear results are DNA fingerprinting, ribotyping, amplified 16S ribosomal DNA restriction analysis (ARDRA), DNA probes, sequence analysis of *fimA* genes and REP-PCR amplicons analysis (Henssge et al. 2011).

As indicated in the discussion of specific points, this study, employing five primer sets (some of them novel) resulted in similar findings to other studies looking at the microbiota of dentine caries. Munson et al. (2004) analysing the microbiota of five carious dentine samples found 95 taxa whn applying two primer sets and culture analysis. They found three taxa to be detectable in all five samples and found *O.*

profusa and *P. acidifaciens* to be predominant in culture analysis, but very few clones could be identified as these two species. In this study detecting a total of 228 taxa were detected of which only a single taxon, *S. mitis/pneumonia*e, could be found in all six patient samples. Similar to the above mentioned study, *O. profusa, O. uli* and *P. acnes* were predominant in culture analysis, but none or only few were detected in molecular analysis. These findings are in contrast to those by Chhour et al. (2005) who, analysing ten carious dentine samples using primer pair 331F / 797R, found *O. profusa* and *P. acidifaciens* to make up to 14.4 % and 30 % of the detected bacterial load, respectively. Studies by Tanner et al. (2011) and Kanasi et al. (2010) looked at culture and molecular analysis of the same plaque samples taken from children with and without early childhood caries (ECC). In those children suffering from ECC samples included plaque from cavities in both studies. The culture study reported a strong association of *S. mutans*, but also found diverse microbiota as well as a novel potential pathogen associated with ECC, *Scardovia wiggsiae* (Tanner et al. 2011). Phyla found in the current study using culture were also observed in those two studies, but at 28 taxa streptococci demonstrated a higher diversity compared to the 17 found here. The most abundant taxa (39) belonged to the genus *Prevotella*. A higher diversity by culture compared to clonal analysis was reported and explained by primer bias (Tanner et al. 2011). Kanasi et al. (2010), analysing the same sample pool using molecular techniques reported 139 taxa (and 35 provisional taxa) and found *S. mutans* and *Bifidobacteriaceae* to be significantly associated with ECC. However, this was only discovered using specific PCR primers, indicating that caries sites are highly diverse and that, while important in disease, these may be present only in low proportions. Another explanation could of course be that the universal primers are not specific enough to allow for full detection of those species. In this study *S. mutans* was detected in five of the six patient samples using culture and molecular techniques, albeit in varying frequencies. These results support the theory that it is the combination of some patients having low numbers of this species and that the choice of primers greatly influences rate of detection.

The most prevalent species found in the culture and clonal study were *S. mutans*, *S. wiggsiae*, *V. parvula*, *S. cristatus*, *A. gerensceriae* and *Granulicatella elegans* and *Veillonella* HOT780 (Kanasi et al. 2010; Tanner et al. 2011). Of these only *S. mutans*, *V.*

parvula and *S. cristatus* were also found in this study, possibly highlighting differences between caries in primary and permanent dentition. Kanasi et al. (2010) using molecular methods did not detect either bifidobacteria or other species belonging to the phylum of *Actinobacteria* nor lactobacilli. It is well established that species belonging to the phylum *Actinobacteria* can be found in caries lesions, in fact, a minimum of five taxa belonging to this phylum being detected per patient sample and Kanasi et al. (2010) themselves detected a positive correlation of lactobacilli with ECC. All the above mentioned study findings support the theory that primer choice has a huge effect on the level of detection that can be achieved. In fact, had only primer pair 27F YM / 1492R been used 62 taxa that were detected in this study would have been overlooked. Even using primer pair 27F CM / 1492R that allowed detection of the most taxa not found with any other primer combination would result in the oversight of 51 taxa.

If no culture analysis had been done, none of the 57 *O. profusa* isolates would have been detected and thereby this taxon would have been completely missed. It seems less likely that novel species will be detected using established culture methods, but not impossible, as can be seen in the case of Tanner et al. (2011). In that study, sampling efforts were extensive with carefully chosen media and incubation conditions, which resulted in the cultivation of 45 previously uncultivated taxa and 45 potential novel groups.

3.6 Conclusion

The work described here confirms findings of previous studies in that caries lesions harbour a wide diversity of bacteria and that each individual has a unique microbiota, while at the same time patients can be grouped according to higher taxonomic levels such as genus and family. Although it was hoped the application of novel primers could increase detection of high G+C species, this was not the case and further research in this area may be needed.

Since culture analysis was found to complement molecular analysis, the recommendation from this study would be to not abandon efforts to culture anaerobic organisms from patient samples. It appears only by applying both methods can the full picture of the oral microbiota be revealed. Results furthermore strongly suggest using

several primer pair combinations, since between six and 17 taxa and potentially many novel sequences would have been missed had any one of the primer combinations not been included in this study.

Chapter 4

Molecular characterisation of the bacterial community in dentine caries using 454 pyrosequencing

4 Chapter 4

4.1 Introduction

The PCR / cloning / sequencing approach using the Sanger sequencing method, which was used to amplify the samples in Chapter 3, has greatly expanded our knowledge of the composition of complex bacterial communities but does suffer from some disadvantages. PCR and cloning reactions are known to introduce bias, whilst Sanger sequencing is time consuming and labour intensive, so that relatively few samples can be processed.

As a result, new culture-independent methods for the profiling of microbiota have been developed in recent years and the method of choice to analyse the oral microbiome is pyrosequencing, because of its combination of high throughput and adequate read length. The most widely used platform to date is the Roche 454 system. Pyrosequencing uses emulsion PCR and a bead-based sequencing reaction in individual wells of pico-titer plates. Light signal peaks resulting from a chemiluminescent reaction in the pyrosequencing cycle are detected by a charge coupled device (CCD), and are proportional to the number of each type of incorporated nucleotides, enabling the deduction of the DNA sequence (Mardis 2008). This technique claims to remove biases introduced in the cloning reactions used in the Sanger-based methods.

One of the first studies applying this method to study the oral microbiota was that reported by Keijser et al. (2008). However, species richness was overestimated due to homopolymers (repeated single bases) that can lead to erroneous base calling as well as formation of chimeric sequences that were mistakenly included. Problems associated with data quality and analysis are being addressed and re-evaluated constantly (Quince et al. 2009; Reeder et al. 2009; Schloss et al. 2011; Diaz et al. 2012) and improved read lengths of up to 500 bp as well as refined data analysis tools have resulted in better phylogenetic resolution and increased confidence in study findings.

Consequently, pyrosequencing has been used in an increasing number of studies to evaluate the oral microbiota associated with health and specific disease states. For example, Li et al. (2010) investigated bacterial diversity in endodontic infections, while at the same time directly comparing 454 sequencing results with conventional Sanger capillary sequencing. A 600-fold increase in ‘depth-of-coverage’ was observed when

the two sequencing technologies were compared. The difference of 8 *vs.* 13 phyla and 25 *vs*. 179 genera in Sanger sequencing and pyrosequencing, respectively, led to the conclusion that although overall endodontic infections appear to be comprised of a less diverse microbiota compared to saliva and supragingival plaque, pyrosequencing showed that endodontic infections are more diverse than previously demonstrated (Li et al. 2010).

Another study compared pyrosequencing with the microbial community profiling microarray assay HOMIM and found that both methods were highly correlated at the phylum level and, when comparing the more commonly detected taxa, also at the genus level (Ahn et al. 2011). Further studies have used 454 sequencing to compare microbiota, for example, in periodontitis and health (Griffen et al. 2012), showing distinct community profiles and significant differences at all phylogenetic levels in health and disease. Similar observations were made when the bacterial diversity of acute vs. chronic, asymptomatic root canal infections were compared (Santos et al. 2011). *Fusobacteria*, for example, were more prevalent in acute than chronic infections. Both studies noted a more diverse bacterial community in (acute) disease compared to health/asymptomatic infections and overall findings revealed a greater diversity than shown in previous studies addressing the same issues. A study examining the salivary microbiota of caries active and caries-free subjects found that *Prevotella* species were not just more abundant in disease, but subjects also carried different arrays of species in health and disease (Yang et al. 2012). Whilst this study did not find subjects with caries to contain a more diverse community structure, they did find the salivary microbiome of caries active patients to be more variable compared to the relatively conserved salivary community structure in health.

It appears useful and advisable to first assess the comparability of next generation sequencing methods to conventional and established community profiling methods, before these methods are used by themselves. As was shown in Chapter 3, culture analysis reveals different findings compared to the PCR / cloning / sequencing methodology and complements the molecular techniques. It is not yet clear, if pyrosequencing alone can detect the full diversity and richness. Furthermore, caution is advised, as the sequencing technique has not yet reached its full potential,

particularly in terms of read lengths, while especially the data analysis is far from maturity and consensus in the various laboratories.

4.2 Aims

The aim of this chapter was to revisit the microbiology of the caries lesions assessed in Chapter 3 using pyrosequencing in order to assess detection of bacterial species with regards to taxa and to evaluate if those numbers were comparable to results obtained with Sanger sequencing.

Another key aspect of the 454 analysis was to see if this method of sequencing that avoids any cloning reaction, would result in raised detection of high G+C bacteria.

4.3 Materials and methods

4.3.1 16S rRNA gene PCR

For amplicon library construction, 16S rRNA genes of the extracted DNA from the initial patient sample were amplified using the six barcoded forward primers 27FYM-A-77 - 27FYM-A-82 and one reverse primer 519-R-B (Table 4.1) as is needed for 454 sequencing using the Lib-L emPCR method. Multiplex-Identifier Sequences (MID's) for the multiplexing of samples and template specific regions to amplify a suitable portion (~500 nucleotides in length) of the 16S rRNA gene (including hypervariable regions 1-3) were also included. Three replicate amplification reactions were set up for each sample and unique forward primer/519-R-B set. Reactions were prepared containing 12.5 µl Extensor PCR mastermix (High fidelity Taq polymerase, Thermo Scientific), 2 µl of template, 0.5 µl of each primer (10 µM) and 9.5 µl sterile water. Initial denaturation was at 95 °C for 5 min, followed by 25 cycles of denaturation at 95 °C for 45 sec, annealing at 53 °C for 45 sec, extension at 72 °C for 90 sec and a final extension at 72 °C for 15 sec. A negative no template reaction was set up for every primer set/sample.

4.3.2 Separation and visualisation of DNA and PCR products

PCR products and molecular weight markers (100 bp DNA ladder, BioLabs) were subjected to electrophoresis in 1 % agarose gels as described (section 2.3.8)

4.3.3 PCR product purification

PCR amplicons were pooled and purified using the QIAquick PCR purification kit (Qiagen, Crawley, UK) following the manufacturer's instructions to remove un-used primers and nucleotides. Purified product was eluted in 30 µl 0.1 x TE buffer

4.3.4 Quality control of purified amplicons

Size and purity of purified amplicons were evaluated using the Agilent 2100 Bioanalyzer along with the Agilent DNA 1000 kit (Agilent Technologies, Inc., Wokingham, UK). This was to ensure that no primer dimers were present and that products were of the correct size. To progress to emulsion-PCR and sequencing on the GS-FLX, the DNA had an OD 260/280 ratio of 1.8 or above and was at a concentration of 5 ng / µl or greater.

4.3.5 Quantitation of amplicons

Accurate quantitation of the amplicons with the Quant-iT-Picogreen fluorescent nucleic acid stain (Invitrogen), a fluorometric assay, was performed to determine concentration (ng / µl) of each amplicon. This was converted to molecules per µl and subsequently amplicons for each library were pooled in equimolar concentrations (1 x 10^9 molecules / µl) (Performed at Centre for Haemato-Oncology, Barts Cancer Institute, Queen Mary University of London).

4.3.6 emPCR and unidirectional sequencing

The samples were amplified clonally by emulsion-PCR using the GS emPCR Lib-L Kit. The GS PicoTiterPlate Kit was then used to sequence individual clonally amplified molecules on a Roche 454 GS-FLX Titanium sequencer. (Performed at Centre for Haemato-Oncology, Barts Cancer Institute, Queen Mary University of London).

Table 4.1: Composite fusion primers consisting of specific sequence (blue) along with unique 'barcode' sequences (black) and the Roche GS-FLX-454 Titanium series adaptor sequences (A = red, B = green) for the Lib-L kit emPCR method.

Primer name	Sample	Template specific sequence	Golay barcode	Adaptor A/Adaptor B (lib-l)
27FYM-A-77	K-A	AGAGTTTGATYMTGGCTCAG	TAACTCTGATGC	CCATCTCATCCCTGCGTGTCTCCGACTCAG
27FYM-A-78	K-B	AGAGTTTGATYMTGGCTCAG	TAAGCGCAGCAC	CCATCTCATCCCTGCGTGTCTCCGACTCAG
27FYM-A-79	K-C	AGAGTTTGATYMTGGCTCAG	TACACACATGGC	CCATCTCATCCCTGCGTGTCTCCGACTCAG
27FYM-A-80	K-D	AGAGTTTGATYMTGGCTCAG	TACACGATCTAC	CCATCTCATCCCTGCGTGTCTCCGACTCAG
27FYM-A-81	K-E	AGAGTTTGATYMTGGCTCAG	TACAGATGGCTC	CCATCTCATCCCTGCGTGTCTCCGACTCAG
27FYM-A-82	K-F	AGAGTTTGATYMTGGCTCAG	TACAGTCTCATG	CCATCTCATCCCTGCGTGTCTCCGACTCAG
519-R-B		GWATTACCGCGGCKGCTG		CCTATCCCCTGTGTGCCTTGGCAGTCTCAG

4.3.7 In silico analysis of DNA sequencing results

The sequence data was subjected to the mothur shhh.flows command to de-noise the data. The trim.flows command was used to remove the primer sequences and barcodes, sequences shorter than 350 bp, and sequences with mismatches in barcode, primer and sequences.

The data was de-replicated using unique.seqs and aligned to the silva.bacteria 16S rRNA reference file by means of align.seqs (Pruesse et al. 2007). Sequences that had more than two ambiguous bases, that did not start by position 1044 (97.5 %- tile) or end by position 5710 (2.5 % -tile) were removed using the screen.seqs command. At the same time, a minimum length criterion of 350 bp was set. Any columns with a '-' in every sequence were removed using filter.seqs and any further redundant sequences were removed using unique.seqs again. Pre.cluster was used to merge sequences that were within 1 bp per 100 bp of total sequence length of a more abundant sequence with that sequence.

Chimerae were detected using chimeras.uchime and removed using remove.seqs. The classify.seqs command was used to classify sequences using the HOMD version 10 reference sequence and taxonomy databases. The dist.seqs program calculated uncorrected pairwise distances between aligned DNA sequences and the cluster command was used to assign sequences to OTUs. Following this, a table was created indicating the number of times an OTU was present in each sample using the make.shared command. Because the groups for the different patients contained varying amounts of sequences, all samples were normalised to the size of the smallest sample group (4293 sequences) in a randomised manner using the command sub.sample. The classify.otu command was used to obtain a consensus taxonomy for each OTU at a value of 98.5 % (or 0.015) using the HOMD version 10 reference sequence and taxonomy databases.

The collect.single command was used to calculate the Chao1 richness and the Inverse Simpson diversity index, while the rarefaction.single command was used to compile rarefaction curve data. A table containing the number of sequences, sample coverage, number of observed OTUs and the Inverse Simpson diversity estimate was compiled using the summary.single command.

4.3.8 Statistical analysis

A two tailed Z-Test Calculator for paired comparisons was used with a significance threshold of 0.05.

4.4 Results

The initial steps of de-noising and trimming the data resulted in 35191 sequences. Of the 1627 unique sequences, 27 were found to represent chimeras (1.6 %) and were removed. Following the quality control steps described above, the patient groups consisted of 4345, 5244, 6488, 4527, 4938 and 4293 sequences for patients A, B, C, D, E and F, respectively. After random sub-sampling to normalise sequence numbers, all patient samples consisted of 4293 sequences (see also Table 1, Appendix 2).

4.4.1 Mothur computed statistical analysis

The coverage rates for the sequencing effort undertaken ranged between 97.2 and 99.3 % using the Good's coverage estimator, but estimated coverage rates using CatchAll values indicated much lower values at 25.5 – 71.81 % (Table 4.2). The number of observed sequences in the samples was 67 to 331 taxa, with patients A and E displaying the lowest values and patient B the highest. Values for the Inverse Simpson diversity estimator show a similar spread of 2.66 (A) to 35.53 (B). Chao1 values also gave the lowest estimates for patients A and E, though in this instance values for patient E (121) were lower than for patient A (142.6), while data for patient B resulted in the highest estimate at 486022.

Table 4.2: mothur summary for each set of 454 patient data detailing number of sequences for each subject (nseqs), observed OTUs (S_{obs}), Good's coverage, CatchAll richness estimate, coverage estimate based on CatchAll value, Chao1 richness estimator and the Inverse Simpson diversity index.

Subject	nseqs	S_{obs}	Good's coverage	CatchAll	CatchAll cov. %	Chao1	invsimpson
A	4293	67	0.9935	93.3	71.81	142.60	2.66
B	4293	331	0.9720	1223	27.06	486.22	35.53
C	4293	176	0.9839	690.1	25.50	282.64	13.10
D	4293	287	0.9746	1052	27.28	441.89	18.72
E	4293	77	0.9923	216.8	35.52	121.00	2.72
F	4293	187	0.9844	337.2	55.46	263.24	11.18

4.4.1.1 Rarefaction curves

Rarefaction curves were computed for each patient and are shown in Figure 4.1. All of the curves levelled off to a certain degree with increasing amounts of sequences

included in the analysis. Curves of patient B showed the steepest slope at the beginning with an elevation of more than 80 % within the first 200 sampled sequences. Visual inspection indicated a similar development in initial slope and levelling further on, for curves representing data for patients C and F. In fact, the only difference was that the final count of patient F's OTUs was 6 higher than the final data point of the graph of patient C. Curves for patients A and E were also similar in slope and development, but the curve for patient E indicated a slightly greater number of final OTUs at 77, compared to the 67 of patient A.

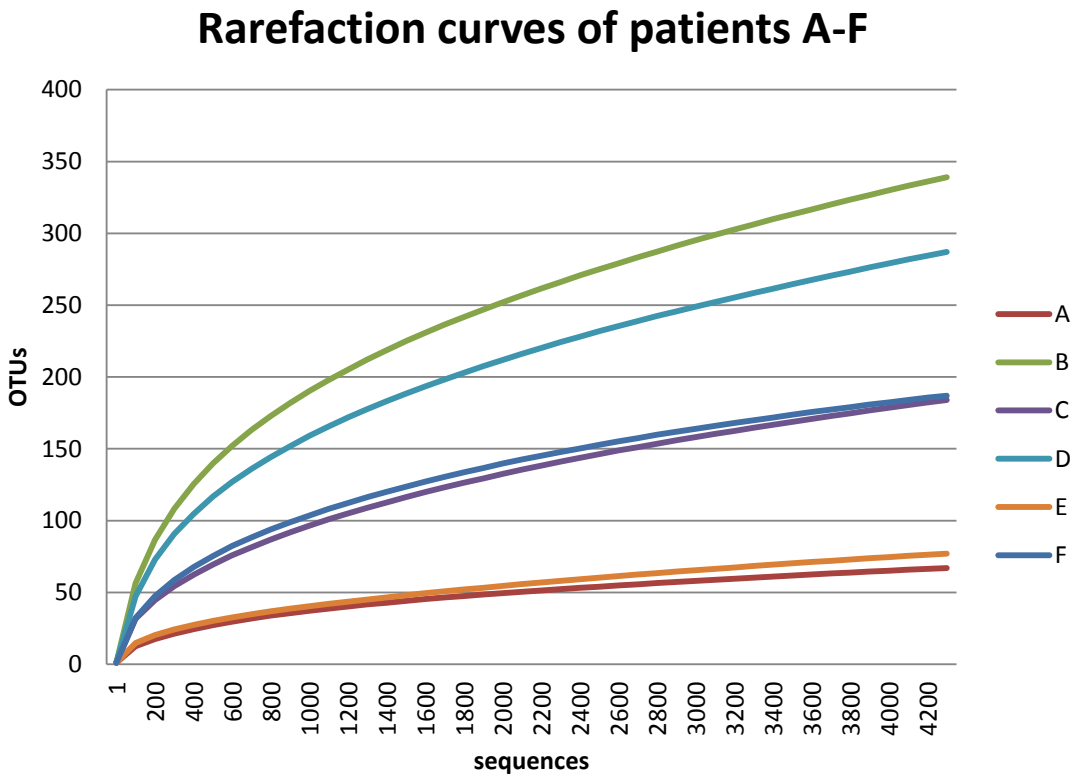


Figure 4.1: Individual rarefaction curves of microbial populations from dentine caries lesions in six patients. Using primers 27F and 519R in 454 sequencing PCR amplicons of at least 350 bp length were used for analysis. The numbers of different OTUs (Operational Taxonomic Unit) in each sample are plotted vs. the number of sampled sequences.

Table 4.3 shows a comparison of OTU based analysis of 454 and Sanger sequencing data using primer set 27F YM and reverse primer 519R and 1492R, respectively (sequencing primer for most Sanger sequences 519R). The number of sequences available from 454 sequencing for analysis is between is between 47.7 and 49.34 times

greater than from Sanger sequencing, while the number of observed OTUs (S_{obs}) is between 4.05 and 7.44 times greater in 454 sequencing. Although the coverage rates for patient A are the closest match – 99 % for the 454 data and 97 % for the Sanger sequencing data, for this patient, the Chao1 estimated species richness shows the greatest discrepancy with the Sanger sequencing data resulting in a prediction of a little over 14 times fewer OTUs compared to 454 sequencing. In all other comparison pairs, the Chao1 value for 454 sequencing is between 1.88 and 8.27 times greater than that for Sanger sequencing, while coverage rates range between 97 and 99 % for 454 sequencing and are much lower at 67 – 88 % for the remaining Sanger-sequenced patient samples. Inverse Simpson diversity indices are higher for most Sanger sequencing samples, with the exception of patients A and C, where Sanger sequencing values are lower by 12 and nearly 17 %, respectively.

Table 4.3: mothur analysis summary for patient data for the 454 data (A-F) as well as the Sanger data (A_Lib2-F_Lib2) detailing number of sequences for each subject (nseqs), observed OTUs (S_{obs}), Good's coverage, Chao1 richness estimator and the Inverse Simpson diversity index.

Subject	nseqs	S_{obs}	Good's coverage	Chao1	invsimpson
A	4293	67	0.99	142.60	2.66
A_Lib2	87	9	0.97	10.00	2.34
B	4293	331	0.97	486.22	35.53
B_Lib2	88	51	0.67	73.56	42.53
C	4293	176	0.98	282.64	13.10
C_Lib2	90	25	0.88	34.17	10.88
D	4293	287	0.97	441.89	18.72
D_Lib2	90	42	0.73	69.60	23.42
E	4293	77	0.99	121.00	2.72
E_Lib2	89	19	0.84	64.50	3.34
F	4293	187	0.98	263.24	11.18
F_Lib2	90	39	0.73	94.20	15.64

4.4.2 Taxonomic assignment

Two methods, namely the mothur version of the "Bayesian" classifier and the HOMD 16S rRNA Sequence Identification tool, were used to assign sequences to the HOMD 16S rRNA RefSeq Version 10.1, a reference taxonomic database.

Analysis of the 25758 sequences resulted in identification of 264 taxa at species level using the Bayesian classifier, representing 11 phyla: *Firmicutes*, *Proteobacteria*, *Actinobacteria*, *Bacteroidetes*, *Fusobacteria*, TM7, *Spirochetes*, *Synergistetes, Chloroflexi,* SR1 and one unclassified (Table 1, Appendix 2).

The distribution of sequences among phyla is shown in Figure 4.2.

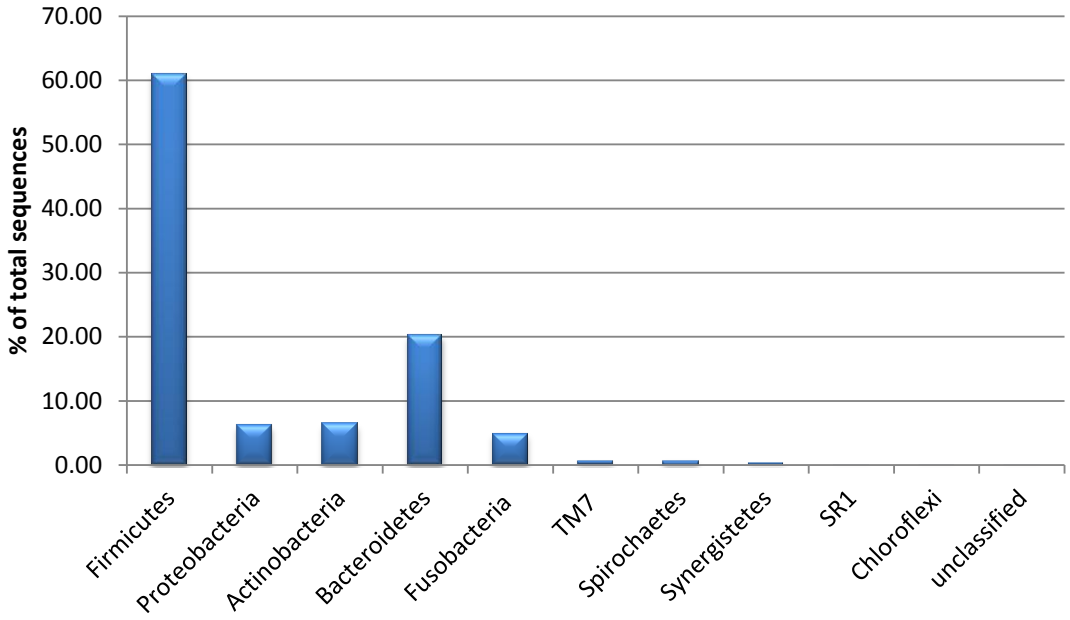


Figure 4.2: Distribution of sequences among bacterial phyla detected with 454 sequencing using primers 27F YM/519R for initial amplification.

Taxa belonging to the phylum *Firmicutes* represented by far the highest proportion at just over 60 % of the total sequences. *Bacteroidetes* were the next most prevalent with just over 20 %, followed by *Actinobacteria* at 6.36 %. Detection of *Proteobacteria* and *Fusobacteria* were at similar levels to those of *Actinobacteria* (6.20 % and 4.83 %, respectively), while detection of all other phyla was well below 1 %.

Proportions of detection using 454 sequencing appeared similar to those observed with Sanger sequencing (Figure 4.3), but were significantly different for *Firmicutes*,

Actinobacteria and *Fusobacteria*. Detection rates of the major phyla (> 1 %) using the mothur analysis was most similar to that of library 2 (27F YM / 1492R) using Sanger sequencing for *Firmicutes*, *Bacteroidetes* and *Fusobacteria*, i. e. there was no significant statistical difference between detection rates. The same was true for rates of detection of *Proteobacteria,* which were closest to that of library 1 (27F CM / 1492R).

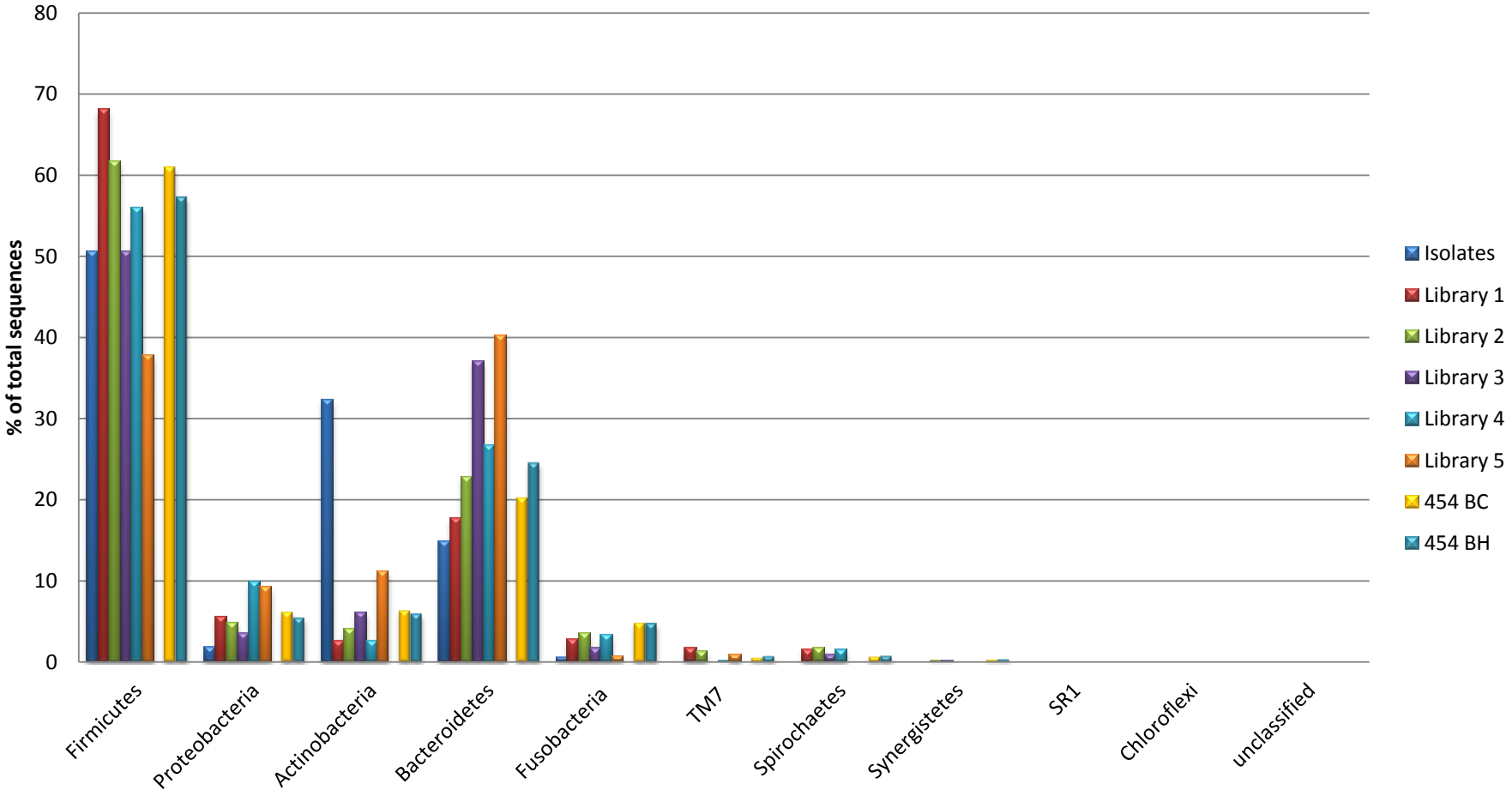


Figure 4.3: Diagram of sequences among bacterial phyla detected using Sanger sequencing (Isolates and Libraries 1 – 5) and 454 sequencing analysed using the Bayesian classifier in the mothur suite (454 BC) and BLAST search via the HOMD 16S rRNA identification tool (454 BH). Primers used for Sanger sequencing: Isolates - 27F CM / 1492R, Library 1 – 27F CM / 1492R, Library 2 – 27F YM / 1492R, Library 3 – 39F / 1387R, Library 4 - 39F / 1492R, Library 5 – 61F / 1387R. Primers used for 454 sequencing: 27F YM / 519R.

The 454 sequencing approach using the mothur analysis resulted in a value of 6.36 % for the detection of *Actinobacteria*, which appeared to be most similar to that of library 3 (39F / 1387R, detection of *Actinobacteria* 6.11 %) using Sanger sequencing as can be seen in Figure 4.3, and was not significantly different ($p = 0.8181$).

The detection rate of *Actinobacteria* using culture was 32.22 % - nearly 3 times higher and therefore significant ($p < 0.05$) compared to the highest detection rate using Sanger sequencing. Comparing the proportional culture analysis detection rates to 454 sequencing using Bayesian classifier and BLAST via HOMD showed that culture detected 5.06 and 5.37 times as many *Actinobacteria*, respectively, which was also statistically significant ($p < 0.05$).

454 sequencing using mothur analysis resulted in the detection of 87 named genera. Sixty-five of the genera detected using 454 sequencing were also found using the Sanger sequencing, whilst 25 were only detected using 454 sequencing identification. In comparison, four named genera were found only with Sanger sequencing and not 454. Of the 25 genera found only with 454 sequencing most (17 of the 25) were represented with five or less sequences. The genus *Corynebacterium*, which was represented with the most sequences (100 sequences, nearly three times the amount of the genus in second place (*Bulleida*) with 39 sequences), still only represented 0.38 % of all sequences.

The analysis using mothur also returned 34 unclassified OTUs represented by 3685 sequences (14.30 %), which could potentially present novel phylotypes. Nine of the 34 unclassified OTUs belonged to the same genera/families in which confirmed novel phylotypes were detected using Sanger sequencing (data not shown).

Of the 35191 sequences run through the HOMD 16S rRNA sequence identification tool 2990 showed homology of < 98.5 % to the database, indicating they could be novel phylotypes. This represents 8.5 % of the total sequences and those 2990 sequences showed closest homology to 165 oral taxa (76.7 – 98.4 % homology), representing a total of 110 genera (data not shown).

As with the results for Sanger sequencing, *Streptococcus mutans* and lactobacilli, the organisms traditionally associated with dental caries were seen in nearly all samples.

Similar to Sanger sequencing results, a highly diverse community was seen, including numerous representatives of the genus *Prevotella* (34 OTUs), *Lactobacillus* (11 OTUs), *Streptococcus* (16 OTUs), *P. alactolyticus* and *Fusobacteria* (7 OTUs). The five most abundant species were *Lactobacillus gasseri*, a group of unclassified streptococci, *S. mutans*, *P. alactolyticus* and *P. denticola*; together representing 40.65 % of all detected sequences. Three of the five most abundant species were the same as for Sanger sequencing results (*L. gasseri*, *S. mutans* and *P. denticola*), but proportions were significantly different ($p < 0.05$).

Differences in the composition of the microbiota in samples from different subjects were observed as seen in the Sanger sequencing analysis and the same trends of numbers (proportions/presence/absence) were observed for the 454 data. For example, both sequencing methods showed samples of patients A and E had significantly higher proportion of lactobacilli than the other patients, with few other *Firmicutes* observed in patient A, whereas the sample from patient E showed a high proportion of *S. mutans* and *Veillonella parvula* ($p < 0.05$). Using Sanger sequencing of patient samples B, C, D and F resulted in the detection of virtually no lactobacilli, but a wide range of other *Firmicutes* (60 - 113 taxa); however, 454 sequencing did not show such a high diversity. Patients B, C, D and F revealed 28 - 63 taxa, compared to 21 and 20 taxa for patient A and E, respectively. Contrasting to that, but in keeping with findings from Sanger sequencing, only few *Prevotella* species were detected in patient samples A and E (0 - 0.32 % of sequences), while significantly more sequences of patient B (14.2 %), C (43.5 %), D (37.1 %) and F (11.9 %) were made up of *Prevotella.*

Other examples highlighted in Chapter 3, such as a high detection rate of *P. alactolyticus* and *Atopobium* OT416 in patient C and more taxa belonging to the genus *Porphyromonas* being detected in patients B and D (3 versus 0 - 1 in the other patients) were confirmed. In addition, detection of the majority, namely 90.3 %, of all detected sequences identified as *O. profusa* belonged to sample F ($p < 0.05$), similar to findings of Sanger sequencing results.

However, a few exceptions to the rule were discovered. Using Sanger sequencing, *Actinomyces* sequences were not significantly differently distributed amongst patients, with the exception of patient C, where no *Actinomyces* sequences were detected,

which was significant compared to all other patients. The 454 data revealed a slightly different pattern with patient B revealing significantly more sequences belonging to this taxon (5.75 %, $p < 0.05$). Patient F exhibiting a little over half as many sequences (3.47 %) and was significantly different not only to patients A, C, D and E (0.61, 0.65, 0.98 and 0.58 %, respectively, $p < 0.05$), but also patient B ($p < 0.05$).

The numbers of detected sequences for the genera *Olsenella*, *Eubacterium* [11][G-6] and *Haemophilus* stood out in that the proportions of sequences for Sanger sequencing were significantly higher than those for the 454 sequencing. *Olsenella,* for example, made up between 0 – 0.68 % of sequences, while using Sanger sequencing patient A was the only patient with *Olsenella* species below the detection level and the remaining patients showed significantly higher levels of 0.19 – 6.85 % ($p < 0.05$). With both analysis methods patient F showed the highest detection levels at 0.68 % using 454 sequencing and 6.85 % using Sanger sequencing, though this was significant only in the case of 454 data.

In the case of *Eubacterium* [11][G-6] sequences were only detected in patient B at a level of 0.07 % using 454 sequencing, while Sanger sequencing resulted in significantly higher detection of sequences in patients B, C and D ranging from 0.37 – 0.56 %.

Haemophilus sequences were not detectable in patient C with either sequencing method, but for all other patients levels ranged from 0.02 – 0.63 % using 454 sequencing, while Sanger sequencing showed proportions of 0.19 – 4.44 %. Pair-wise comparisons of total sequences for this genus, as well as pair-wise comparisons between each patient, showed significantly higher detection rates for all Sanger data, except for patient D, where the detection rate was not significant ($p = 0.08186$).

Comparisons of 454 results with species found using culture analysis show that ten, possibly 11, species were found that were not detected using 454 sequencing (Table 4.4). For one species this is not clear, since *Neisseria flava/mucosa/pharyngis/sicca* could not be assigned a single phylotype with certainty. If the isolate is *N. pharyngis* it has not been detected with 454, the other three possibilities have been detected with 454 sequencing.

Table 4.4: Species that were detected using culture techniques and Sanger sequencing, but not 454 sequencing

Species	Phylum	Number of isolates
***Olsenella uli* Oral Taxon 038**	*Actinomyces*	19[4]
***Propionibacterium acnes* Oral Taxon 530**	*Actinomyces*	1
***Capnocytophaga* sp. Oral Taxon 336**	*Bacteroidetes*	1[4]
***Sphaerocytophaga* S3 sp. Oral Taxon 337**	*Bacteroidetes*	1[4]
***Oribacterium* sp. Oral Taxon 102**	*Firmicutes*	1[4]
***Selenomonas infelix* Oral Taxon 639**	*Firmicutes*	1[4]
***Staphylococcus epidermidis* Oral Taxon 601**	*Firmicutes*	1
Staphylococcus hominis	*Firmicutes*	1
***Veillonella* sp. Oral Taxon 158**	*Firmicutes*	1
***Veillonellaceae* oral taxon 131**	*Firmicutes*	1[4]
Neisseria flava / mucosa / pharyngis / sicca	*Proteobacteria*	1[4]

[4] Species also detected using Sanger sequencing

4.5 Discussion

In this chapter, the microbial load of the six patient samples collected from soft dentine caries lesions for culture and molecular analysis in Chapter 3 were analysed using one set of primers (27F YM / 519R) and 454 sequencing. A wide diversity of bacteria was detected in comparisons between the patients' samples. Distribution of taxa was similar to that of Sanger sequencing results using five different primer sets and culture analysis.

The aim of this chapter was to review the microbiology of the caries lesion using one primer pair and assess detection of bacterial species with regards to taxa as well as numbers found per taxon and to evaluate if those numbers were comparable to results obtained with Sanger sequencing. A search of the literature revealed that no 454 studies targeting the 16S rRNA gene to assess the diversity of caries microbiota have yet been published, much less the direct comparison of Sanger sequencing results using various primer pairs and culture analysis and 454 sequencing results.

Since one of the aims of Chapter 3, the increased detection of high G+C species in the molecular analysis, was not achieved, a focal point of the 454 analysis was to see if this method of sequencing that avoids any cloning reaction, would result in raised detection of these microorganisms. As it turned out, 454 sequencing did not result in detection rates comparable to those found in culture analysis.

The bootstrap value for the mothur analysis was set to 80 as recommended by Schloss et al. (2009). Therefore, sequences designated as unclassified could represent potentially novel sequences or sequences with homology greater than 98.5% to sequences in the database, which resulted in ambiguous taxon assignment at species level. These sequences would need full-length sequencing of the 16S rRNA gene to resolve this uncertainty. Unclassified genera represent novel genera, as there cannot be any ambiguity at genus level. Therefore, it is known that eight novel genera have been detected, but it is not known if any sequences representing novel species were detected, due to the potential ambiguity in identification, for example, of *Streptococcus* and *Neisseria* species (Hanage et al. 2005; Do et al. 2009).

The likelihood that some novel phylotypes have been detected is high nonetheless, since nine genera/families matched those of the novel phylotypes found with Sanger

sequencing. This suggests that 454 sequencing may have identified the same novel phylotypes as Sanger sequencing. It is likely that further novel species have been detected due to number of sequences and the corresponding likelihood of finding rare/novel species. On the other hand, it is also likely that a high proportion of those 3685 sequences represent phylotypes that cannot be assigned a species name with certainty due to ambiguity at the identification level chosen. It is therefore probable that more novel sequences were detected using 454 sequencing compared to Sanger sequencing. However, this would be expected due to the number of sequences screened and significance with regards to numbers can only be established once 16S rRNA genes of potential novel phylotypes have been fully sequenced.

Screening through the HOMD 16S rRNA identification tool showed 2990 sequences below the cut-off threshold of 98.5% for species assignment, indicating these are novel. However, it is unlikely that all of these sequences are considered novel when full length sequencing of the 16S rRNA has been completed. In addition, it is not known how many potentially novel sequences were found per taxon level or how many potentially novel sequences at each taxon level were detected for each patient. The work involved for this detailed analysis was beyond the scope of this study and is one of the challenges that high throughput sequencing has introduced to current data analysis.

The difficulty associated with obtaining full length sequences of potential novel sequences is one of the added disadvantages of 454 compared to Sanger sequencing, since with Sanger sequencing clones and isolates can be stored long term and retrieval of samples and subsequent full length sequencing is relatively easy. If full length sequencing was required from mothur analysis results, new and very specific forward primers for the target sequence would need to be designed and applied to the original patient sample (together with a universal reverse primer). The primer design alone might be difficult due to the short sequence obtained from 454 sequencing. Following successful amplification of the target, the PCR product would then have to be cloned, clones partially sequenced to confirm presence of the target species and subsequent full length sequencing of the target performed.

As is the nature of 454 sequencing, nearly 11 times more raw sequences and nearly eight times more quality controlled and normalised sequences were obtained compared to Sanger sequencing results. These figures would be even higher if it was taken into consideration that for the Sanger sequencing five clone and one culture library were used and the 454 sequencing used only one primer set. Even though the number of sequences was far higher in this instance, coverage rates for both sequencing approaches appeared good, with Sanger sequencing resulting in values of 96 - 98 % across libraries and values for individual libraries, especially patient A, reaching 97 – 99 %, while the 454 sequencing resulted in coverage rates of 97 – 99 % across all patients. It is evident that use of 454 sequencing did increase coverage compared to Sanger sequencing, especially in patients with a diverse microbiota, such as patients B and D. In addition, 454 sequencing and mothur analysis resulted in a much more even distribution of coverage, whereas Sanger sequencing results of coverage for individual libraries varied from as low as 61 % to 99 %. This clearly demonstrates that a higher sequencing effort for samples such as patient B are needed, while samples such as patient A gave with satisfactory results due to a comparatively low incidence of taxa. Since these attributes of patient samples are unknown prior to data analysis 454 sequencing having such a high return of sequences with comparatively low workload is a great advantage of this type of sequencing method.

However, a study by Diaz et al. (2012) argued that coverage estimates for mock communities using the Good's coverage estimator is not a sufficient measure of richness coverage. Because this estimate is based on singletons it does not seem to be suitable for less evenly distributed communities, as is the case in this study as evidenced by Inverse Simpson diversity indices and a comparison of CatchAll coverage estimates compared to those for Good's coverage. In the cited study Good's coverage estimates ranged from 99 – 99.9 %, while CatchAll results indicated a coverage of merely 57 – 82 %. Similar to findings in the study of Diaz et al. (2012), coverage estimates for CatchAll in the current study are in stark contrast to those of Good's coverage, indicating that a much higher sequencing depth is still needed. Solely the patient A sample had a reasonably high coverage value of nearly 72 % but the remaining samples ranged from 25.5 - 55.5 % only according to CatchAll estimates.

Diaz et al. (2012) showed that CatchAll reaches a relative stability ~3000-5000 sequences, which they argue makes this estimator reliable, even if a greater sequencing effort is indicated. The amount of subsampled species in this study falls exactly within the cited range and it is therefore probable that CatchAll coverage estimates reflect the true coverage achieved better than the calculated Good's coverage.

Pyrosequencing also helped enhance knowledge of microbiota in dentine caries, in that a number of genera were found, which were not seen with Sanger sequencing. Most of those sequences were detected in low abundances and may represent part of the rare biosphere (Pedros-Alio 2012). This can be explained by the sheer amount of sequences generated by 454 sequencing, whereby the chance of detecting a rare species is increased. Indeed, two sequences belonging to the phylum SR1 and one sequence belonging to the phylum *Chloroflexi* were detected that have previously been described as rare (Keijser et al. 2008). Diaz et al. (2012) argued that OTUs representing singletons should be eliminated, since it was found that in communities with known numbers of OTUs 454 sequencing and following analysis generated more OTUs than expected. However, findings from this study show that OTUs from the rare biosphere would thus be missed and steps to eliminate or include OTUs in the analysis have to be given careful consideration (Reeder et al. 2009; Schloss et al. 2011).

The distribution of phyla found with 454 sequencing appeared to be in the range of that detected using the clone libraries and Sanger sequencing. For example, 454 sequencing resulted in the detection of just over 60 % *Firmicutes*, while Sanger sequencing detection rate varied between 38 – 68 % for the different libraries; however, detection rates of *Firmicutes*, *Actinobacteria* and *Fusobacteria* were statistically significantly different. At the same time, the Z-test calculator for paired comparisons confirms the value for detection of *Firmicutes* is not significantly different to that of library 2 using Sanger sequencing. Statistical analysis furthermore confirmed that proportions of phyla detected using 454 sequencing often are most similar to those of libraries 1 or 2 (27F CM or YM / 1492R) as would be expected, since forward primer 27F YM was used in the initial amplification for the 454 analysis.

The only exception to this rule is represented by the detection rate of *Actinobacteria*, which appeared to be closest to that of library 3 (39F / 1387R), and was indeed not significantly different from this library. Apart from culture, libraries using reverse primer 1387R achieved highest detection rates of *Actinobacteria* using Sanger sequencing. The fact that 454 sequencing using primer 27F YM detected proportions similar to those of libraries using reverse primer 1387R could indicate that detection of *Actinobacteria* was slightly improved using this method. Since the primer combination 61F / 1387R resulted in the highest detection rate using molecular methods it might be interesting to repeat the 454 sequencing using 61F as the forward primer. This, however, might introduce different biases in itself, since it was found that despite the mismatch to streptococci being removed, a bias against this genus remained (see Chapter 3 section 3.4.1).

One hypothesis raised in Chapter 3 was the possibility of the introduction of bias when cloning was used; more specifically, that preferential ligation could skew results. Since there is no cloning involved in 454 similar results to those of Chapter 3, which did involve cloning, would indicate that any bias introduced would unlikely be due to cloning or that any bias introduced in cloning reactions would at least be negligible. The target of improving the detection rate of high G+C phylum *Actinobacteria* was not achieved in Chapter 3. Since the detection rate of *Actinobacteria* using 454 sequencing at 6.36 % is in the range of that found using the five different primer combinations using molecular analysis (2.6 – 11.1 %) it can be concluded, at least in this instance, that cloning is not to blame for the 'underrepresentation' of *Actinobacteria*.

In fact, as was discussed in Chapter 3, it is becoming doubtful that *Actinobacteria* are truly underrepresented. It appears that the explanation of *Actinobacteria* appearing as overrepresented in culture is probable. The most likely explanation for this is the fact that many species to date cannot be grown in culture, as has been discussed elsewhere (Wade 2002; Vartoukian et al. 2010). Therefore the proportion of *Actinobacteria* appears raised in comparison to molecular analysis, which can detect those uncultivable species, whilst seemingly lowering the proportion of high G+C species that readily grow on media and culture conditions used in the laboratory as standard. Indeed, in Chapter 3, *O. profusa* was not detected in any of the clone libraries although it was found using culture techniques. A similar outcome was

observed with regards to *O. uli*, where 19 out of the 21 sequences were detected using culture.

Pyrosequencing and subsequent mothur analysis using the Bayesian classifier resulted in 31 sequences being identified as *O. profusa* (three in patient D, 28 in patient F), three sequences identified as *O. uli* (patient C) and a further three sequences identified as *Olsenella* sp. Oral Taxon 807 (two in patient B, one in patient D). Analysis of the 454 data using a BLAST search on HOMD resulted in 41 sequences being identified as *O. profusa* (five in patient D, 36 in patient F), five as *O. uli* (four in patient C, one in patient D), seven as *Olsenella* sp. Oral Taxon 807 (six in patient B, one in patient D) and one sequence as *Olsenella* sp. Oral Taxon 809 (patient F). The HOMD 16S rRNA identification tool reported only the sequence of *Olsenella* sp. Oral Taxon 809 as having less than 98.5 % homology to the one in the database (96.2 %), indicating that it might be by a potential novel phylotype. The Bayesian classifier also returned one sequence designated as unclassified, which might indicate a novel sequence. The discrepancy of total number of sequences identified as belonging to the genus *Olsenella* is down to the fact that mothur uses a lower number of sequences, namely the normalised sequences, whereas all de-noised and trimmed sequences were used for the HOMD analysis. However, the proportion of *Olsenella* for both methods is very nearly the same at ~0.15 % (Mothur 0.1474 % vs. HOMD 0.1534 %) and this low number lends further support to the theory that *Actinobacteria* are actually overrepresented in culture study due to the many as yet-uncultivable species and molecular analysis gives a more accurate reflection of true proportions. This also proves that the primers do amplify at least this genus, and detection rates of the Sanger sequencing results do reproduce proportions of what is in the sample.

In contrast to this, a study attempting the full genome sequencing of the high G+C organism *Actinoplanes* sp. SE50/110 found a high number of gaps (Schwientek et al. 2011). Mapping of the sequencing results on a reference gene cluster sequence revealed a fragmentation into 30 contiguous sequences of different lengths. The gaps between these sequences were characterised by extremely low read coverage which strongly correlated with the G + C content in the gap regions in a negative manner and the gap-sequences contained strong stem-loop structures which hindered the amplification of these sequences during the emulsion PCR (Schwientek et al. 2011). It

is not clear if the high G+C content in 16S rRNA and its adjoining regions would have such a strong effect as it does on whole genome sequencing, but it should be considered a source of continued potential bias and an evaluation of the extend of this bias using mock communities may be useful.

A further suggestion for future work taken from the above study is the inclusion of the additive, emPCR Additive (Roche, 454 Life Sciences), in the emulsion PCR that exerts inhibitory effects on the self-annealing ability of single stranded DNA molecules. This could potentially improve detection of high G+C organisms, just as additives mentioned in previous chapters did in conventional PCR amplification reactions (Schwientek et al. 2011).

A comparison of 454 results to culture results obtained in Chapter 3 showed several species from various phyla were not detected using 454 sequencing. Seven of these potentially 11 species were also found with Sanger sequencing (Table 4.4) and of those seven found using Sanger sequencing, five were found using primer 27F (CM or YM), indicating that 454 sequencing did not detect some sequences that were found by culture and molecular means using Sanger sequencing. This could hint at a possible bias of 454 sequencing. It might be interesting to apply different primers to these samples using 454 sequencing to see if this bias can thus be removed, especially to detect those species that were not detected by Sanger sequencing.

Only one study using pyrosequencing to examine differences in plaque from healthy individuals and individuals with different stages of caries experience has been published to date. However, this study employed a metagenomic approach and the taxonomic classification based on 16S rRNA genes made up a relatively small subset of the total sequences (Belda-Ferre et al. 2012). Furthermore, comparisons to 16S cloning based studies were made at class level, while a comparison of health versus disease status was made at genus level. It is questionable that the phylogenetic resolution is sufficient to draw meaningful conclusions as it has been shown that closely related species can have very different relationships with health and disease states (Wade 2011). This notion is furthermore supported by a study examining the microbiome of caries-active and healthy individuals using 454 sequencing, where findings underscored the necessity of species-level resolution for caries prognosis, since caries-

active and healthy subjects carried different arrays of *Prevotella* species (Yang et al. 2012).

In previously published studies detection of *Veillonella* spp. was reported in most patients at all stages ranging from intact to deep dentine cavities, but no significance could be drawn in relation to any stage of caries lesion progression (Aas et al. 2008; Gross et al. 2010; Lima et al. 2011), showing that the significance of the detection of certain species and genera is still not clear. When Belda-Ferre et al. (2012) made comparisons at species level between health and caries status, however, *Veillonella parvula* was found in caries-active and caries-free patients. Only the metagenomic analysis showed that different strains were present in health and disease, since the *Veillonella* found in caries-active individuals contained genes that *Veillonella* in caries-free subjects did not, which could indicate that different genes are involved in pathogenesis (Belda-Ferre et al. 2012). Though the work presented here and the study described above had different aims, it may nonetheless be useful to keep in mind that 16S based analysis can obscure such associations and that for species found in health and disease analysis, such as a metagenomic one, may further resolve health and disease associations.

Pyrosequencing is a powerful tool to achieve a great sequencing coverage for any given habitat. But the advantages of achievable coverage depth, the potential of detecting members of the rare biosphere, the reduced time and cost involved to obtain comparatively many sequences, and the increased likelihood of detecting true clusterings/associations of species amongst each other and in relation to disease/habitat have to be carefully weighed with the disadvantages and limitations of this method. Diaz et al. (2012) showed that detection of species abundance is subject to empirical bias introduced through methods for DNA isolation and amplification. Though, with regards to amplification bias, an experiment comparing the influence of various primer pairs on detection might help give an indication if presently used amplification primers could be amended or exchanged for more appropriate ones. Diaz et al. (2012) argue, for example, that streptococci were overrepresented in their 454 study as well as in a study based on 16S rRNA gene amplification and subsequent Sanger sequencing by Kroes et al. (1999). However, shown in Chapter 3, the application of different primers shows an apparent under-representation of

streptococci when primer 61F is used. Following the claim made above, it may however be the case that primer 61F actually gives a truer representation of streptococci than the other primers and experiments using this primer in conjunction with a mock community as done by Diaz et al. (2012) could show if this hypothesis can be proven correct.

Further disadvantages of pyrosequencing are the short sequence reads, many of which may fall through quality control and which can make subsequent primer design needed for full length sequencing challenging. The huge amount of sequences can also be problematic in that not every single sequence can be reviewed individually and sequences may be eliminated in quality control steps that are actually representing good quality sequence reads of true (as yet undetected) species.

The huge discrepancy of the Good's coverage estimator and the CatchAll estimated coverage indicate that a much higher coverage is needed. Although further taxa and potentially novel sequences have been discovered using this method, it is not clear if the undetected sequences would change proportions of species detected with Sanger and 454 sequencing so far and if this would have a significant impact on the thinking about caries diagnosis and treatment. Looking solely at results obtained from the six patient samples using Sanger and 454 sequencing, which are of a comparable nature, it would seem Sanger sequencing is as effective in representing true proportions of species for each patient, but only experiments with higher coverage could confirm this. As mentioned previously, it would be very interesting to evaluate in further experiments with these samples if different primers may also influence detection of various species.

4.6 Conclusion

The results obtained from re-analysing the microbiology of the caries lesions assessed in Chapter 3 using 454 pyrosequencing instead of Sanger sequencing show that a greater diversity of species was detected, whilst at the same time, a superior coverage depth in comparison to findings from Chapter 3 was achieved.

The two central questions were: 1) If the pyrosequencing method was capable of detecting the full diversity and species richness alone; 2) Whether this method, which

does not rely on cloning reactions, would result in raised detection of high G+C bacteria.

Findings from Chapter 3, such as each individual has a unique microbiota and that subjects can be grouped according to higher taxonomic levels, were confirmed. However, even though pyrosequencing resulted in the detection of an overall greater diversity, some genera were only found with the Sanger sequencing method, illustrating that pyrosequencing at this coverage depth using this primer pair does not detect the full diversity and species richness by itself.

Furthermore, the overall detection rate of *Actinobacteria* was lower compared to not just culture analysis, but also to library 5 using Sanger sequencing. This showed that although detection was enhanced compared to the commonly used primer combination 27F / 1492R, further improvements in the technique, such as using different primers or using additives, need to be evaluated.

Chapter 5

Development of a FISH protocol for the visualisation of bacteria in caries lesions

5 Chapter 5

5.1 Introduction

In the previous chapters, the microbiota associated with dentine caries was confirmed to be highly diverse. What is not clear is whether all of the bacteria present are playing a role in the infection. The location of individual organisms within lesions is likely to be important, particularly at the advancing front where those bacteria may be actively involved in tissue degradation and at the surface of the biofilm, which has easy access to nutrients from the mouth. It would also be of interest to know if different species co-localise or coaggregate and thereby possibly form symbiotic networks helping in the disease progression.

The extent to which bacteria progress in the advancing lesion is still under debate (Kidd et al. 1993; Banerjee et al. 2002) and illustration with the help of fluorescence *in situ* hybridisation (FISH) and confocal laser scanning microscopy (CLSM) would be invaluable to inspect which species invade to which level and if/how the different species co-localise and therefore possibly interact.

FISH has been shown to be a useful tool for the analysis of the spatial organization of (developing) biofilms *in vitro* as well as *in vivo* (Thurnheer et al. 2004; Diaz et al. 2006). This method uses oligonucleotide probes labelled with a fluorescent dye that bind the 16S rRNA molecule of its target (Amann et al. 1995; Moter et al. 2000). FISH probes can be designed like primers to bind to a broad range of species, such as the eubacterial probe EUB338, or more specifically to genera or individual species (Amann et al. 1995). Targeted microbes can be visualised subsequently using either an epifluorescence microscope or a CLSM. Using FISH, quantitative and qualitative experiments enabling examinations of proportions in natural habitats as well as co-localisations and interactions of species are made possible. A further advantage of this method is that it can also be applied to uncultivable species, enabling characterization of species without the need for cultivation (Schramm et al. 2002). The only pre-requisite is the knowledge of the 16S rRNA sequence to design probes to a level of required specificity.

In Chapter 3, the most prevalent organisms detected with both molecular and culture analysis combined were *Lactobacillus*, *Prevotella* and *Streptococcus* species. Because

of difficulties in validating the newly designed *Lactobacillus* probe candidates due to the cell wall of this genus, *Prevotella*, the second highest-ranking genus, was chosen for FISH probe development in this study. Members of the family *Bifidobacteriaceae*, that are known to have a mismatch with primers 27F CM and 27F YM, ranked further down with regards to detected numbers, but FISH probes targeting *Bifidobacteriaceae* were designed nonetheless, since these are most likely to be underrepresented as a knock on effect from the primer bias. Furthermore, bifidobacteria have been shown recently to be caries-associated and further research into their role in the caries process has been recommended (Beighton et al. 2010). Because *Streptococcus mutans* is strongly associated with dental caries, a probe from a published paper by Thurnheer et al. (2001) was also included in this study.

5.2 Aims

The overall aim was the design and validation of labelled oligonucleotide probes to facilitate analysis of the prevalence and localisation of bacterial species through the advancing dentine caries lesion using FISH and CLSM. The aim was to develop a multiplex FISH method by which the proportion of specific targeted genera/species in excavated tooth debris from a caries lesion could be evaluated in comparison to total bacterial count, which would be marked with the eubacterial probe. The secondary aim was develop a protocol for (multiplex) FISH directly on extracted, bisected teeth.

5.3 Materials and methods

5.3.1 Reference strains for FISH experiments

Table 5.1: Bacterial strains used for FISH and FISH probe validation and respective culture media. BA – blood agar, FAA – fastidious anaerobe agar, NAM - N-acetylmuramic acid.

Species	Strain no.	Media
Actinomyces naeslundii	NCTC 10301	FAA
Streptococcus sanguinis	ATCC 10556^{T}	FAA
Atopobium parvulum	ATCC 33793	FAA
Bifidobacterium dentium	NCTC 11816	FAA
Olsenella uli	ATCC 49627	FAA
Parascardovia denticolens	CCUG 35728	FAA
Scardovia inopinata	CCUG 35729	FAA
Atopobium rimae	ATCC 49626	FAA
Propionibacterium acnes	ATCC 6922	FAA
Rothia dentocariosa	DSM 43762	BA
Neisseria subflava	CCUG 23930	BA
Aggregatibacter actinomycetemcomitans	ATCC 43718	FAA
Lactobacillus gasseri	DSM 20243	FAA
Streptococcus mutans	NCTC 10449	FAA
Prevotella denticola	ATCC 35308	FAA
Porphyromonas gingivalis	ATCC 33277	FAA
Fusobacterium nucleatum ss nucleatum	ATCC 25586	FAA
Prevotella tannerae	ATCC 51259	FAA
Prevotella nigrescens	NCTC 9336	FAA
Prevotella intermedia	ATCC 25611	FAA
Bacteroides zoogleoformans	ATCC 33285	FAA
Tannerella forsythia	ATCC 43037	FAA + NAM
Capnocytophaga gingivalis	ATCC 33624	FAA
Lactobacillus vaginalis	DSM 5837	FAA
Lactobacillus rhamnosus	DSM 20021	FAA
Lactobacillus crispatus	DSM 20584	FAA
Enterococcus faecalis	ATCC 19433	FAA

Species	Strain no.	Media
Granulicatella adiacens	DSM 9848	FAA
Staphylococcus aureus	NCTC 6715	FAA
Slackia exigua	ATCC 700122	FAA

5.3.2 Bacterial culture

Strains (Table 5.1) were grown on FAA / 5 % horse blood (supplemented with 0.001 % (w/v) N-acetylmuramic acid (NAM, Sigma) in the case of *Tannerella forsythia*) in an anaerobic workstation at 37 °C. Both *Rothia dentocariosa* and *Neisseria subflava* were grown on BA aerobically at 37 °C. All other strains were grown on FAA.

5.3.3 Protocol for FISH on bacterial smears

Actinomyces naeslundii, Bifidobacterium dentium, Prevotella denticola, Streptococcus mutans and Slackia exigua were grown as described. Colonies were harvested from agar plates using a sterile inoculation loop into sterile PBS to an approximated density of 10^6 cells per ml. A 10 µl aliquot was pipetted onto each 6 mm-diameter well on gelatine-coated (0.075 % gelatine, 0.01 % $CrK(SO_4)_2$ in H_2O heated to 70 °C) Shandon Multi-Spot microscope slides (Thermo Scientific) and allowed to dry for approximately 30 min at 37 °C. Thirty µl of 1:1 PBS/EtOH solution was added to the well and the slide incubated in a humid chamber at 4 °C for 2 h, followed by two washes with PBS. Slides were then dipped successively in 50 %, 80 % and 96 % ethanol for 3 min each and subsequently air-dried in a horizontal staining jar. Slides were either further processed immediately or wrapped in tin foil and stored at -70 °C.

For hybridisation, 8 µl hybridisation buffer (18 % v/v 5M NaCl, 2 % v/v 1 M Tris-HCl (pH 8.0) and 0.1 % v/v of 10 % SDS) + 1 µl (16 µM) probe (EUB338 or NONEUB as negative control, Table 5.6) were added to the well and incubated in a humid chamber at 50 °C for 2 h. Following hybridisation, the slides were washed with washing buffer (4.2 % v/v 5M NaCl, 2 % v/v 1M Tris-HCl (pH 8.0), 1 % v/v 0.5M EDTA and 0.1 % v/v of 10 % SDS) by first letting it drip over the surface and then immersing it in 50 ml pre-heated washing buffer for 15 min at 52 °C. Finally, the slides were air dried, Vectashield (Vector Laboratories Inc., Peterborough, UK) applied, covers applied and the edges sealed with nail varnish, ready for microscopy.

5.3.4 Shortened FISH protocol

In a shortened version of the above protocol, incubation times for permeabilisation and hybridisation were each reduced from 2h to 30 min.

5.3.5 Development of a protocol for FISH on bisected teeth

Further patient samples were collected under ethical approval (Reference 08/H0810/61). Freshly extracted teeth with suitable carious lesions were placed into a sterile universal container. The teeth were superficially cleaned by immersion in PBS and patted dry with tissue. Samples were stored at 4 °C unless handled. Teeth were bisected using a Labcut 1010 hard tissue microtomes (Agar Scientific, Stansted, UK) with a Diamond wafering blade (high concentration) XL 12205, (Benetec Limited, London, UK) and processed as soon as possible. Teeth were positioned in a humid chamber on a plastic grid cut plane facing up. For the permeabilisation of cells at least 30 µl of 1:1 PBS/EtOH solution was added to each tooth half, ensuring that the whole plane, but especially the lesion, was covered, and the tooth incubated in a humid chamber at 4 °C for 2 h, followed by two washes with PBS. Dehydration was performed as described previously (section 5.3.3).

For the hybridisation, 16 µl hybridisation buffer and 1 µl (16 µM) probe were added to the caries lesion and the protocol followed as outlined in section 5.3.3. Finally, the tooth halves were air dried, Vectashield applied to a clean slide and the tooth immersed in the Vectashield upside down. Microscopy was performed as soon as possible. Teeth were stored at 4 °C if it was not possible to perform microscopy immediately. It was always ensured that the cut plane of the tooth was in contact with Vectashield, protecting the tooth from desiccation.

5.3.6 FISH on hydroxyapatite (HA) discs (adapted from (Thurnheer et al. 2004))

Hydroxyapatite discs were washed twice in PBS and incubated for 15 min at 50 °C in hybridisation buffer without probe. *A. naeslundii* and *P. denticola* were grown as described in section 5.3.2. From these plates, a suspension of turbidity equivalent to McFarland standard 4 (Fisher Scientific, UK) was prepared in sterile saline. Discs were inserted into individual wells of a 24-well advanced TC™ treated multiwall plate (Greiner Bio-one, Stonehouse, UK), covered with 1.4 ml pre-reduced LB broth and 200

µl bacterial suspension. The plates were covered with lids and incubated anaerobically at 37 °C for i) 6 h and ii) overnight (Table 5.2). Following incubation, HA discs were removed using sterile tweezers and washed by dipping them in 2 ml saline three times for 1 min each time. Fifty µl of 1:1 PBS/EtOH solution was added to each disc and incubated in a humid chamber at 4 °C for 30 min. This was followed by two washes with PBS and dehydration of cells by successive submersion in 50 %, 80 % and 96 % ethanol for 3 min each. Discs were subsequently air dried lying flat, 40 µl probe mix (hybridisation buffer + EUB338 or NONEUB probe) added and incubated for 1 h at 50 °C in a humid chamber. A post-hybridisation wash was performed by submerging the discs for 15 min in 7 ml washing buffer in individual pots warmed to 52 °C. Discs were subjected to a final rinse in dH_2O and air dried. Vectashield was applied to a clean microscopic slide and the disc immersed in the Vectashield bacteria-side down, ready for microscopy.

Table 5.2: Combination of strains, incubation periods and FISH probes used in the experiment of FISH on HA discs

Strain	Incubation time	FISH probe
A. naeslundii	6 h	EUB338
A. naeslundii	6 h	NONEUB
A. naeslundii	o/n	EUB338
A. naeslundii	o/n	NONEUB
P. denticola	6 h	EUB338
P. denticola	6 h	NONEUB
P. denticola	o/n	EUB338
P. denticola	o/n	NONEUB

5.3.7 FISH on scoops taken from a bisected caries lesion

Teeth with a carious lesion previously determined for removal were extracted, cleaned and bisected as described in section 5.3.5. Using a sterile excavator, scoops were taken from the infected area of the lesion that was leathery to touch. Scoops were diluted in 50 µl dH_2O and the suspension vortex mixed for 1 min. The solution was spun down for 5 sec in a Biofuge centrifuge at 9.447g and the supernatant used following the protocol described in section 5.3.3. Positive control slides were prepared using a colony of *B.*

dentium as described (section 5.3.3). A shortened protocol was followed in that incubation for permeabilisation and hybridisation lasted for 1.5 h and 1 h, respectively.

5.3.8 Design of FISH probes

Oligonucleotide probes targeting 16S rRNA were designed for the family *Bifidobacteriaceae* and the genera *Lactobacillus* and *Prevotella*.

Probes for each genus or family were designed by visual inspection of a 16S rRNA gene sequence alignment of all sequences of the particular genus detected in the patient sample, reference sequences downloaded from the HOMD and NCBI databases together with sequences from related genera from HOMD and/or NCBI.

Criteria for the probe design consisted of: an exact match to the target group, at least three mismatched bases to other phylotypes (ideally located at the 3' end), and (ideally) a brightness class of I to IV indicative of *Escherichia coli* ribosome probe accessibility of > 40 % and a consensus accessibility map for prokaryotes (Fuchs et al. 1998; Behrens et al. 2003). It was aimed to create probes between 18 - 22 nt long with a GC content of 50 % ±10 % and a melting temperature of T_m 52-58 °C. It was also desirable to have at least one G or C base within the last five bases from the 3' end to achieve a GC clamp. More than 3 G or C residues in a row were avoided.

Since multiple probes would be used simultaneously in multi-FISH experiments, probes potentially targeting the same bacteria were designed to bind to different parts of the 16S rRNA molecule. This was also in case some probes proved not to work *in vitro* due to structural constraints of the folded protein.

Probe specificity was confirmed *in silico* by interrogation of the Ribosomal Database Project-II 16S rRNA database (Maidak et al. 2001) and BLAST search of the GenBank nucleotide database. Probes were synthesised, modified with Cy3 for the validation process and one of three fluorophores at the 5' end – Alexa Fluor 488, Cy3 or Cy5 for sample analysis (Table 5.6). Their excitation and emission spectra were sufficiently distinct as to allow multiple probes (one with each of the three fluorochromes) to be used together in multi-FISH experiments without the risk of crossover. The sequence for probe Mut590 (*S. mutans*) was taken from a publication by Thurnheer et al. (2001) and included in the validation process.

Table 5.3: *Bifidobacteriaceae* oligonucleotide probe candidates for FISH. The narrow range panel representing target organisms, while the broad range panel represents distantly related species.

Probe	Sequence	Length	G+C %	Tm	Brightness class (Fuchs et al. 1998)	Brightness class (Behrens et al. 2003)
Bif698	Cy3-TTC CAC CGT TAC ACC GGG AA	20	55	58.4	III	II/III
Bif128	Cy3-GCA TTA CCA CCC GTT TCC	18	55.56	55.7	III	II/IV
Bif837	Cy3-ACA TCC AGC ATC CAC CGT T	19	52.63	56.5	III/VI	III
Bif1452	Cy3-CTG CCC ACT TTC ATG ACT	18	50	53.4	III	-

Narrow range panel	**Broad range panel**
Parascardovia denticolens	*Neisseria subflava*
Bifidobacterium dentium	*Aggregatibacter actinomycetemcomitans*
Scardovia inopinata	*Lactobacillus gasseri*
Atopobium rimae	*Streptococcus mutans*
Propionibacterium acnes	*Prevotella denticola*
Olsenella uli	*Porphyromonas gingivalis*
Rothia dentocariosa	*Fusobacterium nucleatum ss nucleatum*

Table 5.4: *Prevotella* oligonucleotide probe candidates for FISH. The narrow range panel representing target organisms, while the broad range panel represents distantly related species.

Probe	Sequence	Length	G+C %	Tm	Brightness class (Fuchs et al. 1998)	Brightness class (Behrens et al. 2003)
Prev781	Cy3-ATC CAT CGT TTA CCG TGC G	19	52.63	56.5	II/III	II/III
Prev282	Cy3-TCT CAG AAC CCC TAC CGA	18	55.56	55.7	IV	III/I
Prev512	Cy3-CAC GGA ATT AGC CGG TCC	18	61.11	57.9	III	II/IV
Prev1528	Cy3-GTC ACG GAC TTC AGG CAC	18	61.11	57.9	II	III
Prev734	Cy3-ATA CCC GCA CCT TCG AGC TT	20	55	58.4	IV	IV

Narrow range panel

Prevotella denticola

Prevotella tannerae

Prevotella intermedia

Prevotella nigrescens

Bacteroides zoogleoformans

Tannerella forsythia

Porphyromonas gingivalis

Capnocytophaga gingivalis

Broad range panel

Neisseria subflava

Aggregatibacter actinomycetemcomitans

Lactobacillus gasseri

Streptococcus mutans

Bifidobacterium dentium

Olsenella uli

Fusobacterium nucleatum ss nucleatum

Table 5.5: *Lactobacillus* oligonucleotide probe candidates for FISH. The narrow range panel representing target organisms, while the broad range panel represents distantly related species.

Probe	Sequence	Length	G+C %	Tm	Brightness class (Fuchs et al. 1998)	Brightness class (Behrens et al. 2003)
Lacto379	Cy3-CAT CAG ACT TGC GTC CAT TGT G	22	50	58.4	IV/II	-
Lacto223	Cy3-GGT CCA TCC AAG AGT GAT AGC	21	52.38	58.4	IV	IV
Lacto648	Cy3-TTT CCG ATG CGC TTC CTC	18	55.56	55.7	VI	VI
Lacto637	Cy3-CTT CCT CGG TTA AGC CGA	18	55.56	55.7	IV	VI

Narrow range panel	**Broad range panel**
Lactobacillus gasseri	*Neisseria subflava*
Lactobacillus vaginalis	*Aggregatibacter actinomycetemcomitans*
Lactobacillus rhamnosus	*Bifidobacterium dentium*
Lactobacillus crispatus	*Olsenella uli*
Enterococcus faecalis	*Prevotella denticola*
Granulicatella adiacens	*Porphyromonas gingivalis*
Streptococcus mutans	*Fusobacterium nucleatum ss nucleatum*
Staphylococcus aureus	

5.3.9 Validation of FISH probes and optimisation of hybridisation conditions

The probes were validated *in vitro* against a narrow and broad-range panel of bacteria (Table 5.1) specific for each set of probes (Table 5.3 to Table 5.5). The narrow range panel includes members of the targeted genus or family, whereas the broad range panel consists of unrelated oral species that should not be bound by the designed probes.

The protocol detailed in section 5.3.3 was followed. Optimal conditions for probe hybridisation stringency were determined by varying the concentration of formamide (J.T. Baker, London, UK) in the hybridisation buffer at 46 °C. In brief, the buffer included 18 % v/v 5M NaCl, 2 % v/v 1M Tris-HCl (pH 8.0), 0.1 % v/v of 10 % SDS and formamide concentrations of 0, 10, 20, 30 or 40 %. For the hybridisation 8 µl hybridisation buffer + 0.5 µl (16 µM) probe were added to each well and incubated in a humid chamber at 46 °C for 2 h. To avoid photobleaching of the probe, slides were kept covered in darkness. Following hybridisation slides were washed with washing buffer by first letting it drip over the slide and then immersing it in 50 ml pre-heated washing buffer for 15 min at 48 °C. Finally, the slides were air dried, Vectashield applied, coverslip placed and the edges sealed with nail varnish, ready for microscopy.

5.3.10 Specificity of probes

After validation and optimisation of individual probes, probes Bif698, Bif128 and Prev282, Prev734, Mut590 and EUB338 were ordered with their respective fluorochromes as described in Table 5.6 and specificity tested on mixed species samples at a formamide concentration of 10 % as described in Table 5.7. It was ascertained that fluorescent emission was not due to 'bleed-through' caused by lasers of a different wavelength.

Table 5.6: FISH oligonucleotide probes with their respective 5'fluorophores and optimal formamide concentrations (N/D – not determined)

FISH probe	Sequence (5'-3')	5' fluorophore	Length	G+C %	Brightness class (Fuchs et al. 1998)	Brightness class (Behrens et al. 2003)	Optimal formamide concentration
Prev282	TCT CAG AAC CCC TAC CGA	Alexa Fluor 488	18	55.56	IV	III/I	0 – 10 %
Prev734	ATA CCC GCA CCT TCG AGC TT	Cy5	20	55	IV	IV	0 – 10 %
Bif698	TTC CAC CGT TAC ACC GGG AA	Alexa Fluor 488	20	55	IV	IV	10 %
Bif128	GCA TTA CCA CCC GTT TCC	Cy5	18	55.56	III	II/IV	0 – 20 %
Mut590	ACT CCA GAC TTT CCT GAC	Cy5	18	50	N/D	N/D	0 – 10 %
EUB338	GCT GCC TCC CGT AGG AGT	Cy3	18	66.67	N/D	N/D	N/D
NONEUB	ACT CCT ACG GGA GGC AGC	Cy3	18	66.67	N/D	N/D	N/D

Table 5.7: Species and species mixtures and applied FISH probes to test for non-specific cross-reaction of probes

Species	Probe			
B. dentium / S. mutans	Mut590	Bif698	Bif128	
B. dentium / P. buccae	Prev282	Prev780	Bif698	Bif128
S. mutans / P. buccae	Mut590	Prev282	Prev734	
B. dentium / P. tannerae	Bif698	Bif128	Prev282	Prev734
S. mutans / P. tannerae	Mut590	Prev282	Prev734	
S. mutans	Mut590			
P. buccae	Prev282			
P. buccae	Prev734			
P. tannerae	Prev282			
P. tannerae	Prev734			
B. dentium	Bif698			
B. dentium	Bif128			
S. mutans	EUB338			
P. buccae	EUB338			
P. tannerae	EUB338			
B. dentium	EUB338			

5.3.11 Final protocol for FISH on tooth scrapings and bisected teeth

Patient samples were collected and processed as described in sections 5.3.5 and 5.3.7. For samples consisting of caries scoops/scrapings, tooth hemi-sections were inspected visually and three areas of infected/affected dentine determined from which scrapings/scoops were taken with a sterile excavator as illustrated in Figure 5.1.

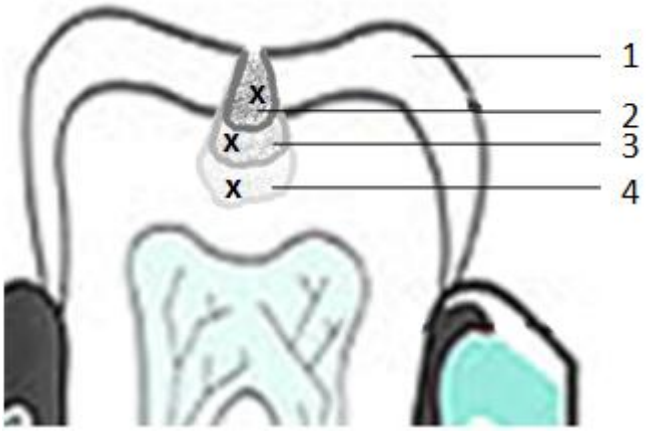


Figure 5.1: Schematic depiction of the zones of carious dentine in active, cavitated, occlusal coronal lesion 1) enamel, 2) outer layer of irreversibly degraded and demineralised infected dentine, 3) inner layer of irreversibly degraded and demineralised infected dentine, 4) affected dentine, X = areas for sample taking. Diagram adapted from http://www.nrch.com.au/oralhealth/oral.htm

Photographs of the tooth hemi-sections were taken using a Nikon D3100 digital SLR camera with a Nikon AF-S 40 mm Micro F/2.8 DX G lens. The areas from which excavated material were taken were marked on the digital image. Single scoops from the infected/affected area were taken and diluted in 155 µl dH_2O and the suspension vortex mixed for 1 min. The solution was pulsed for 5 sec in a Biofuge centrifuge at 9.447g and the supernatant used for FISH on microscope slides.

Thirty microliters (or more if applied directly on teeth) were added to each well/tooth half and samples were incubated in a humid chamber at 4 °C for 2 h. Washing, dehydration and air-drying was carried out as described in sections 5.3.5 and 5.3.7.

In the case of fluorescence oligonucleotide probes applied directly to the tooth 30 µl hybridisation buffer not containing probe was added prior to hybridisation of bisected teeth in a pre-hybridisation step as a blocking agent and incubated for 10 min at room temperature.

Hybridisation was performed as described in sections 5.3.5 and 5.3.7, and samples were incubated at 46 °C for 2 h. The samples were washed at 48 °C and the remaining protocol followed as described (sections 5.3.5 and 5.3.7).

5.3.11.1 Samples 024 and 025

For each scoop from the two teeth, one slide was prepared with five wells each ready for the probes to be applied. Wells were numbered 1-5 left to right. On all slides probe in hybridisation buffer was applied in the following order:

1) Bif698
2) Prev734
3) Mut590
4) EUB338
5) EUB338 + Prev734

5.3.11.2 Samples 026 and 027

Scoops from patient 026 were crushed manually and subsequently vortex-mixed for 5 min. Scoops from patient 027 were vortex-mixed for 5 min only. All suspensions were spun down briefly in a Biofuge centrifuge for 5 sec and supernatant used for slide preparation as described. For each scoop of the two teeth, one slide was prepared with 10 wells each.

Slides were frozen over night at -70 °C and allowed to warm to room temperature the next morning before 10 µl 10 % formamide hybridisation buffer was applied and incubated for 10 min at room temperature. Slides were subsequently rinsed with PBS, excess liquid carefully removed from the slide and probes applied as previously described.

On all slides, probe in hybridisation buffer was applied in the following order:

1) EUB338
2) Bif698 + EUB338
3) Bif128 + EUB338
4) Prev282 + EUB338
5) Prev734 + EUB338
6) MUT590 + EUB338
7) Bif128
8) Prev282
9) Prev734
10) MUT590

As illustrated in Figure Figure 5.2

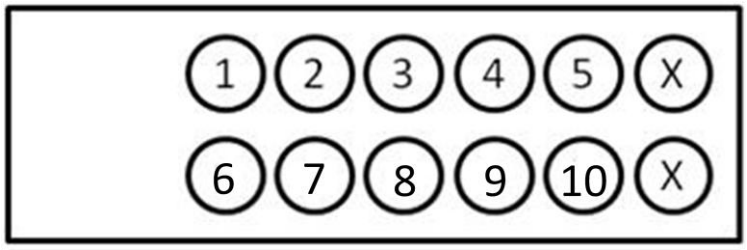


Figure 5.2 Diagram of slide used for FISH

For these samples, the final washing step was repeated in fresh wash buffer.

5.3.11.3 Sample 028

Scoops were taken as described above and each was deposited in an individual, sterile PCR tube containing 20 µl sterile water. The samples were vortex-mixed for 1 min and the supernatant aspirated. All three scoop-samples were used for Gram staining on individual microscope slides. A positive control consisting of supragingival plaque was harvested from a consenting healthy adult using a sterile pipette tip and resuspended in 50 µl sterile water was included.

The other tooth hemi-section was stored at 4 °C over night and FISH performed directly on the tooth as described in section 0. Finally, the tooth sample was mounted on a microscope slide in Vectashield and glued in place with clear nail varnish.

5.3.12 FISH microscopy

Microscopy evaluation of the FISH probes on the labelled narrow and broad range species panels was performed using a Leica SP2 confocal laser scanning system (Leica Microsystems) fitted with argon/argon-krypton laser (operating at 488nm), krypton laser (operating at 568nm) and helion-neon laser (operating at 633nm), with a Leica DMIRE2 inverted microscope (x63 objective).

5.3.13 Gram stain

Scoops from carious lesions diluted in water as described (section 5.3.7) were Gram stained. In brief, an aliquot of the suspension was heat-fixed on a microscopic slide, crystal violet applied for 30 sec and washed off, followed by iodine for 30 sec, which was again washed off and rapid decolourisation with a 1:1 alcohol/acetone solution. The final steps were counterstaining with safranin for 60 sec and drying of the slide.

5.4 Results

5.4.1 FISH on bacterial smears

Microscopy of preliminary experiments with bacterial smears showed a high number of bacteria fluorescing. On most slides, the density was so high that it resembled a lawn of cells. Although the bacterial load on the slide was rather high, the protocol itself did not need further adaptation.

5.4.2 Shortened FISH protocol

The shortened FISH protocol appeared to work as well as the initial protocol when tested on bacterial smears on slides, since fluorescing bacterial cells were readily detected and intensity of fluorescence was comparable to previous experiments.

5.4.3 FISH on bisected teeth

When probe EUB338 was applied to bisected teeth, no bacteria could be detected using CLSM. A repeat of the experiment with fresh teeth and a Tandem scanning confocal microscope (TSM) (Noran Instruments, Middleton, WI, USA) together with a iXon 885 EM-CCD (electron multiplying charge-coupled device) (Andor Technology, Northern Ireland, UK) at 200 x magnification showed a zone of fluorescence congruent with the caries lesion. No individual bacteria could be observed.

5.4.4 FISH on hydroxyapatite (HA) discs (adapted from (Thurnheer et al. 2004))

CLSM performed on each of the eight discs did not result in the detection of any bacterial cells.

5.4.5 FISH on scoops taken from a bisected caries lesion

Microscopy of scoops taken from bisected caries lesions of three patients that were hybridised with EUB338 resulted in detection of fluorescing bacteria in all patient samples (Table 5.8).

Table 5.8: Microscopy results of FISH on scoops taken from bisected caries lesions; +++ 300+ cells, ++ 100-300 cells, + 11—100 cells, +/- 1-10 cells, - 0 cells

sample ID		15	16	19
	rods	+++	+	+
EUB338	cocci	+	+/-	+++
	filaments	+	-	+

Figure 5.3 shows a FISH micrograph of carious dentine in which the bacteria were labelled with probe EUB338-rhodamine. The majority of bacteria that can be observed are rods, with some rather long filamentous cells and some short rods, which might represent diplococci. Only very few cocci can be observed in this picture.

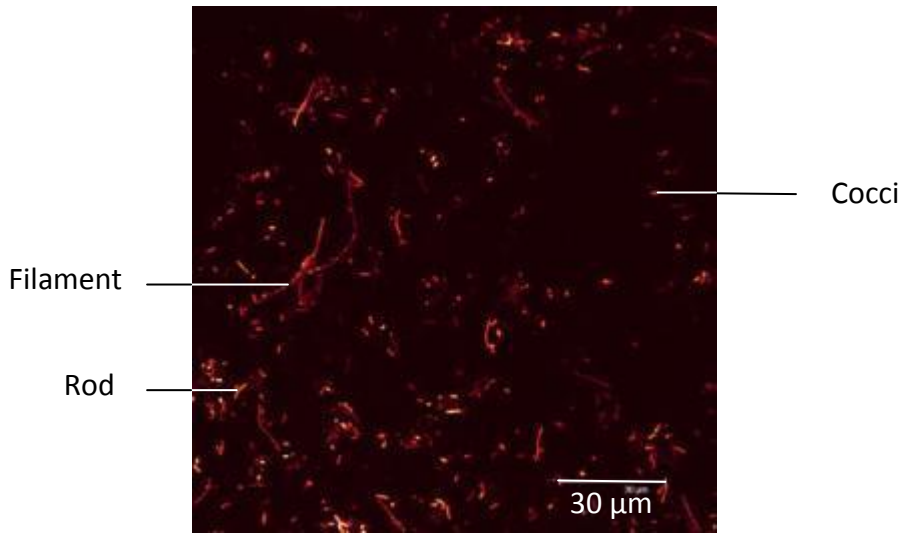


Figure 5.3: Confocal FISH micrograph of a scoop of carious dentine (patient sample 15), using probe EUB338.

Figure 5.4 represents a zoomed in snapshot of Figure 5.3. Again mainly rods and some filaments can be observed.

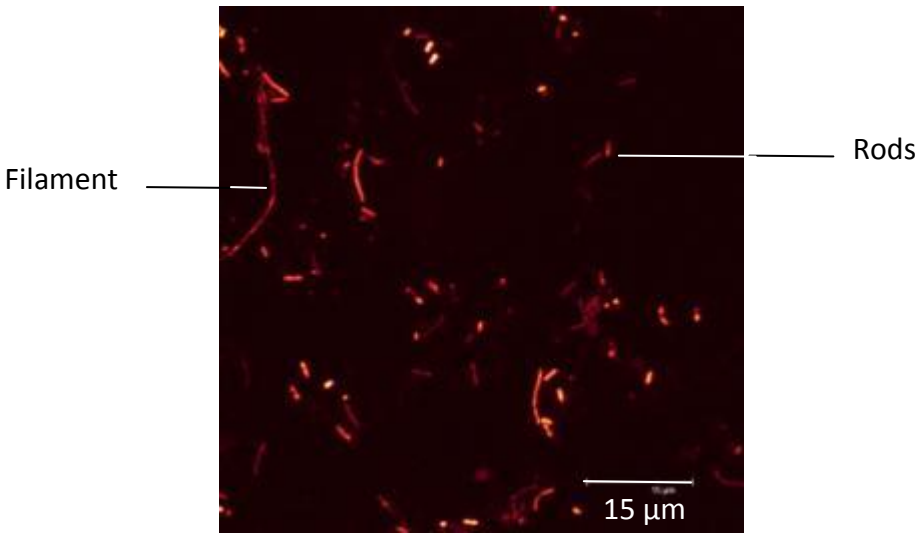


Figure 5.4: Confocal FISH micrograph of a scoop of carious dentine (patient sample 15), using probe EUB338.

Figure 5.5 depicts a very high magnification of the filamentous cells and neighbouring rods. Most rods are straight, while some are slightly curved. The uneven intensity of fluorescence in the filaments represents clustering of RNA, to which the FISH probe binds.

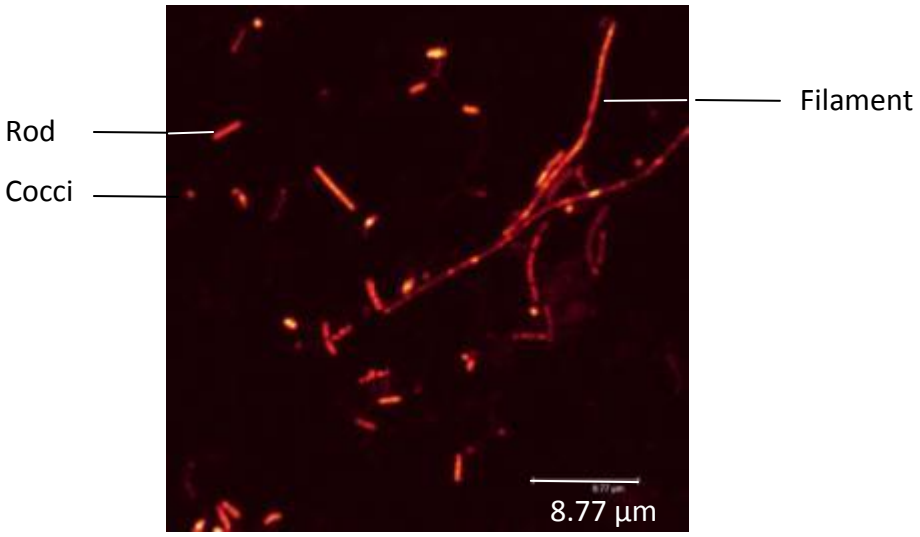


Figure 5.5: Confocal FISH micrograph of a scoop of carious dentine (patient sample 15), using probe EUB338.

Figure 5.6 shows a confocal laser-scanning image of a scoop from the patient 16. Higher magnification compared to patient 15 had to be used to visualize bacteria and to obtain a good quality image. Far fewer bacterial cells compared to the sample of patient 15 were observed. A few curved rods were observed. Some fluorescent points in the upper right quadrant of the image could represent cocci or rods lying vertically to the focus point. No filaments were visible in this patient sample.

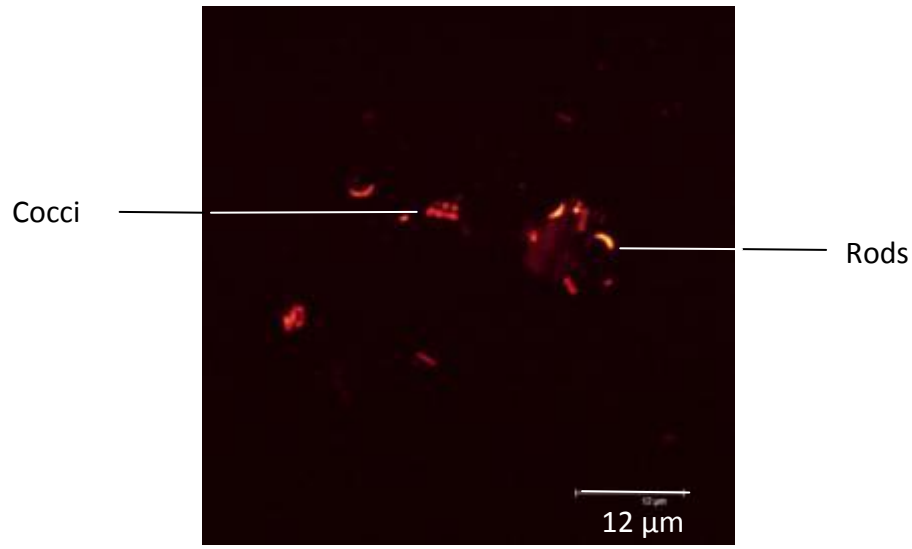


Figure 5.6: Confocal FISH micrograph of a scoop of carious dentine (patient 16), using probe EUB338.

The sample of patient 19 is depicted in Figure 5.7 and consisted mainly of cocci (about 95 %). A few filaments can be seen as well as a few rods.

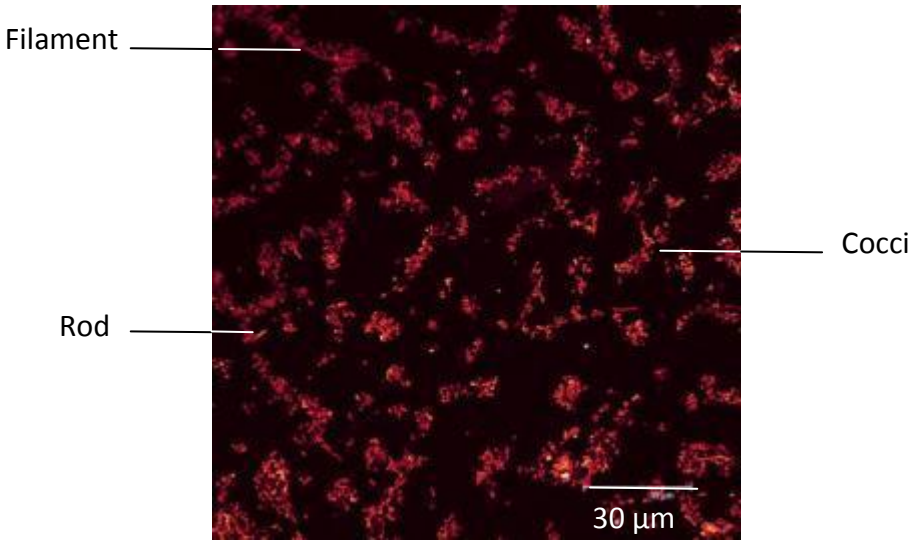


Figure 5.7: Confocal FISH micrograph of a scoop of carious dentine (patient 19), using probe EUB338.

A higher magnification of patient sample 19 in Figure 5.8 shows that many of the cocci appeared as diplococci. Most cocci appeared to be of approximately the same size, but some were smaller than the majority. The filaments appeared mixed with the cocci, but no clear co-aggregation structure such as corncob formations were evident.

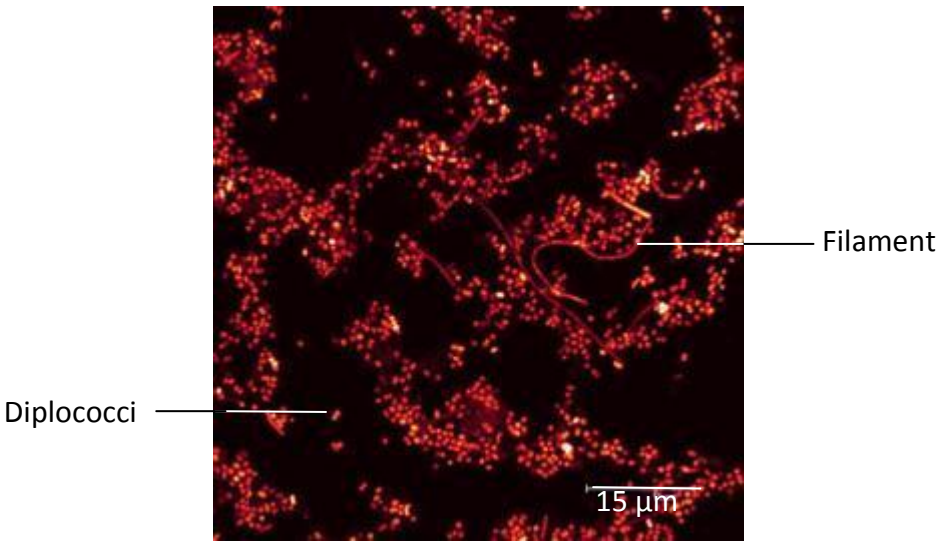


Figure 5.8: Confocal FISH micrograph of a scoop of carious dentine (patient 19). using probe EUB338.

All three patient samples taken were distinctly different from each other. While patient sample 19 appeared to have the highest density in terms of cell numbers, patient sample 15 also exhibited many bacterial cells in the microscopic field at lowest magnification. The proportions of rods and cocci appeared to be reversed in these two patient samples. Patient sample 16, on the other hand, showed very few bacterial cells overall and no filaments.

5.4.6 Validation of FISH probes and optimisation of hybridisation conditions

The following oligonucleotide probes (Table 5.9) were found to be specific to the target organisms in the narrow range panel at the stated formamide concentrations.

Oligonucleotide probes targeting *Lactobacillus* species were not validated. Obtaining good fluorescence in FISH experiments on lactobacilli has been proven difficult due to their cell wall's resistance to probe penetration (Quevedo et al. 2011). In the current study these problems were confirmed and validation of probe Lacto379 resulted in weak fluorescence signal of the narrow panel. Positive control experiments using EUB338 on *L. gasseri* and *L. vaginalis* resulted in fluorescence of about 40 - 60 % of the cells on the slide.

Table 5.9: FISH probe validation detailing probe and its ideal formamide concentration

Probe	% formamide	comments
Bif698	10	
Bif128	0-20	
Bif837	10	*S. inopinata* not as bright as *P. denticolens* and *B. dentium*
Prev282	0-10	*P. nigrescens* slightly weaker at 10 %
Prev512	0-10	At 0 % *B. zoogleoformans* and *T. forsythia* weakly lighting up, at 10 % *P. buccae* and *P. tannerae* slightly weaker signal
Prev781	0-10	At 0 % *B. dentium* weakly lighting up, at 10 % signal for narrow panel less bright than at 0 %
Prev1528	-	*P. tannerae* negative
Prev734	0-10	At both concentrations individual cells were non-specifically lighting up
Mut590	0-10	*F. nucleatum* positive, which was also used by Thurnheer et al. (2001) and was negative

5.4.7 Specificity of probes

Hybridisation of probes with mixtures of species and pure species confirmed specificity of all probes. Good signal intensity was achieved in all cases, although in some cases brightness/probe intake was not uniform. In the case of *B. dentium* mixed with *P. buccae* and probe Prev282, the *P. buccae* cells fluoresced strongly, but shadows could be observed in the background. It is not clear if those shadows were bifidobacteria from the mixture or if these were weakly stained *Prevotella* cells. In the case of *S. mutans* mixed with *P. tannerae* and probe MUT590, individual cells of *P. tannerae* could be seen in the background. The brightness of the *S. mutans* cells was much

stronger than that of the *P. tannerae* cells and cell morphology made it easy to distinguish between the two.

5.4.8 Final protocol for FISH on tooth scrapings and bisected teeth

Two teeth were obtained. Following hemi-section, the two halves were photographed in order to detail from where the sample scoops were taken.

Figure 5.9: Tooth sample 024 with 1 indicating the sample from the innermost region of the lesion, 2 and 3 indicating sampling sites from the leathery and harder, progressing part of the caries lesion, respectively.

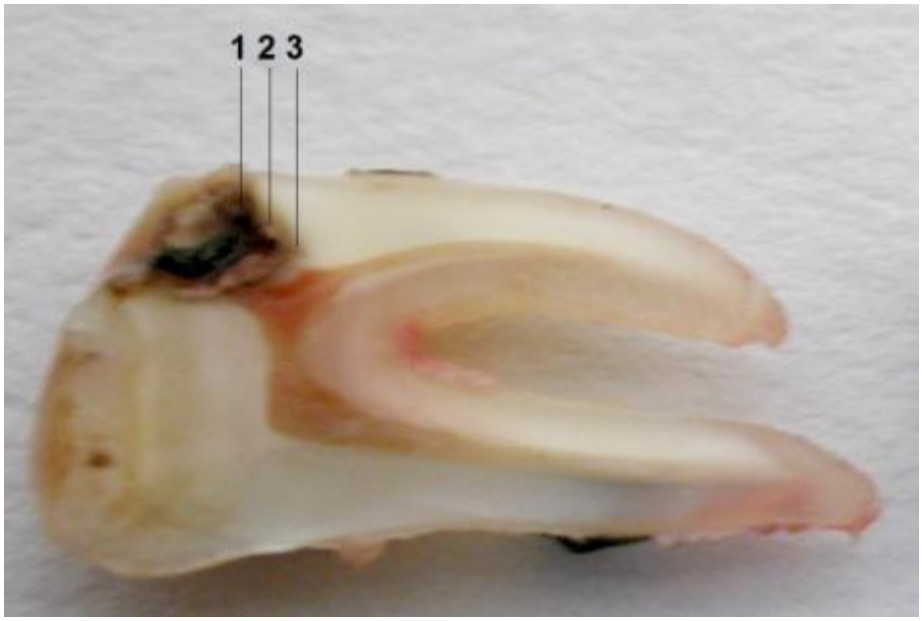


Figure 5.10: Tooth sample 025 with 1 indicating the sample from the innermost region of the lesion, 2 and 3 indicating sampling sites from the leathery and harder, progressing part of the caries lesion, respectively.

CLSM of scoops taken from patient sample 025 demonstrated the presence of many bacterial cells in scoop 1, but detectable numbers decreased in scoop 2 and no bacterial cells could be found in scoop 3 (Table 5.10).

Application of probe Bif698 to sample 025 scoop 1 (Figure 5.11 a) resulted in few cells being observed. Most of those fluorescing were seemingly surrounded by debris. Higher magnification of the same microscopic field showed rods fluorescing that morphologically look like *Bifidobacteria*. Probe Prev734 (Figure 5.11 b) resulted in large clusters of cells being observed that were surrounded by debris. The morphology of fluorescing cells was short rods. Some cells, medium to long rods, in the background were only weakly fluorescing. Fluorescent signal strength of probe MUT590 (Figure 5.11 c) varied over the microscope field. Clusters of cells were observed primarily and individual cells were not recognisable. Most cells showed characteristics of cocci growing in chains, in agreement of what would be expected to be observed using this probe. Some smaller fluorescent points looked like individual cocci. Structures presumed to be debris were also found to be fluorescing. Universal probe EUB338 (Figure 5.11 d) resulted in detection of most cells compared to the other probes. In fact, the microscope field was nearly filled with bacteria, most cells grouping in clusters. Some cloud-like fluorescence in and around clusters was thought to be debris. Another microscopic field showed fewer cells overall and fewer clusters. Short and medium length rods, possibly cocci, some chains of cells and maybe some filaments (it was not clear if they were filaments or chains of cells) dominated both fields.

The application of two probes (EUB338 + Prev734) seen in Figure 5.12 resulted in detection of several small clusters of cells, mostly rods, some cocci, some long rods and some filaments or chains of cells for the universal probe EUB338. For Prev734 it appeared that mainly debris was stained. Individual cells that were stained had lower signal intensity compared to the fluorescence observed with probe EUB338. The overlay of both probe signals showed more blue (Prev734 probe) than red (EUB338 universal probe) indicating that the *Prevotella*-probe bound to the debris and the EUB338 probe more specifically to cells, making all cells appear pink, but fluorescing overall more blue than red. Some cell-like blue-fluorescing signals were more likely to be small debris, to which EUB338 did not bind, but Prev734 did.

Table 5.10: Microscopy results of FISH on scoops taken from three areas of bisected carious lesions; +++ 300+ cells, ++ 100-300 cells, + 10—100 cells, +/- 1-10 cells, - 0 cells

sample ID		024 scoop 1	024 scoop 2	024 scoop 3	025 scoop 1	025 scoop 2	025 scoop 3
	rods	+++	-	-	+++	++	-
EUB338	cocci	++	-	-	+	+/-	-
	filaments	+	-	-	+	+/-	-
Bif698		-	-	-	++	+	-
Prev734		+	-	-	+++	-/+	-
Mut590		+++	-	-	+++	-	-
	rods	+++	-	-	++	-	-
EUB338 /	cocci	+	-	-	-/+	-	-
	filaments	-/+	-	-	-/+	-	-
Prev734		BL[5]	-	-	BL[5]	-	-

[5] BL=Bleed-through of one laser resulting in excitation of the fluorochrome by the wrong laser

To test if the blue colour in the multi-FISH experiments was resulting from autofluorescence a second slide prepared from the same patient sample scoop solution, but not labelled with any probe, was viewed using the microscope under the same operating conditions and no fluorescence of any kind could be observed, strongly suggesting Prev734 had a relatively strong affinity for dental debris.

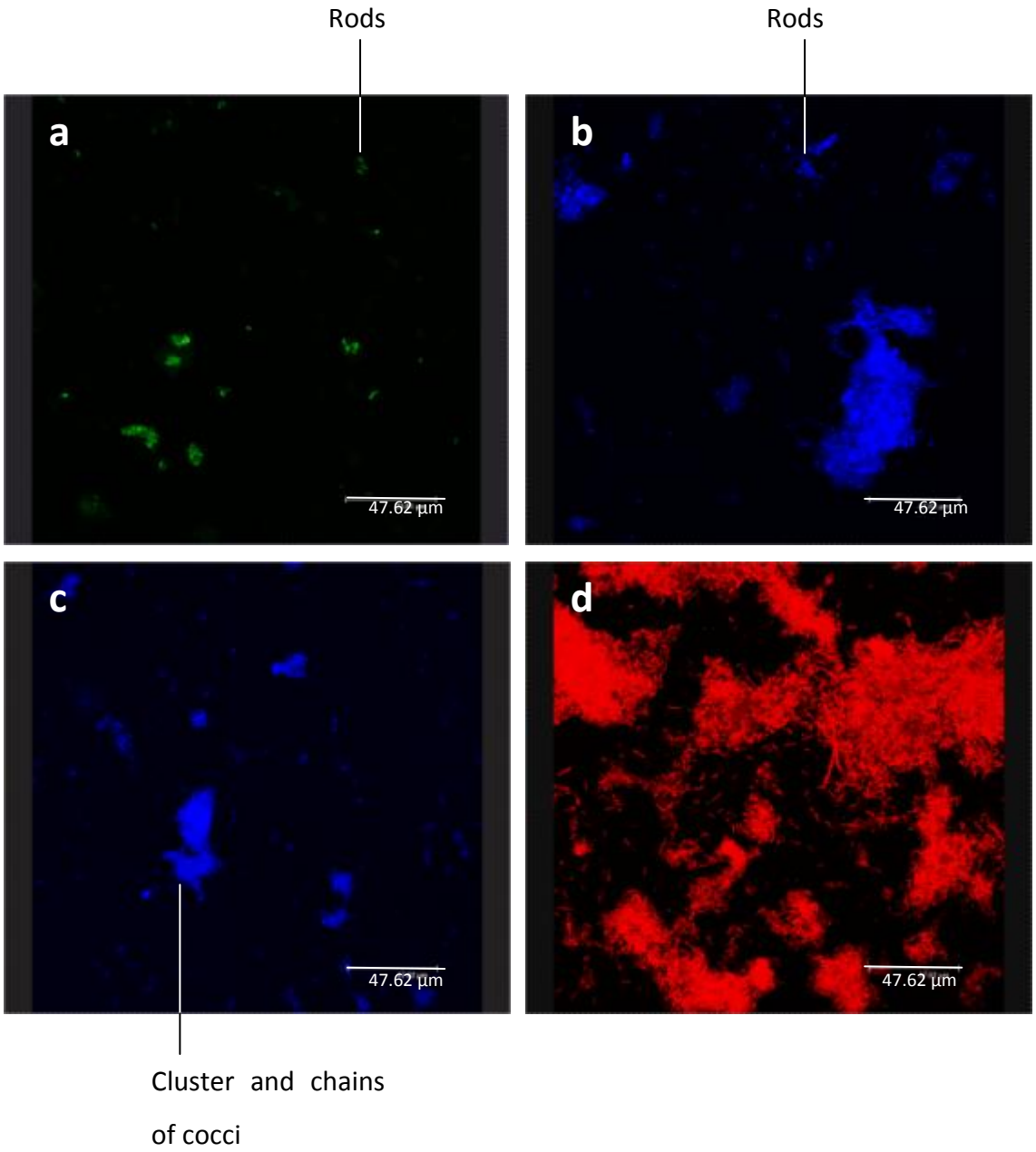


Figure 5.11: Fluorescence *in situ* hybridisation micrographs showing patient sample 025 scoop 1 with probes a) Bif698, b) Prev734, c) MUT590 and d) universal probe EUB338

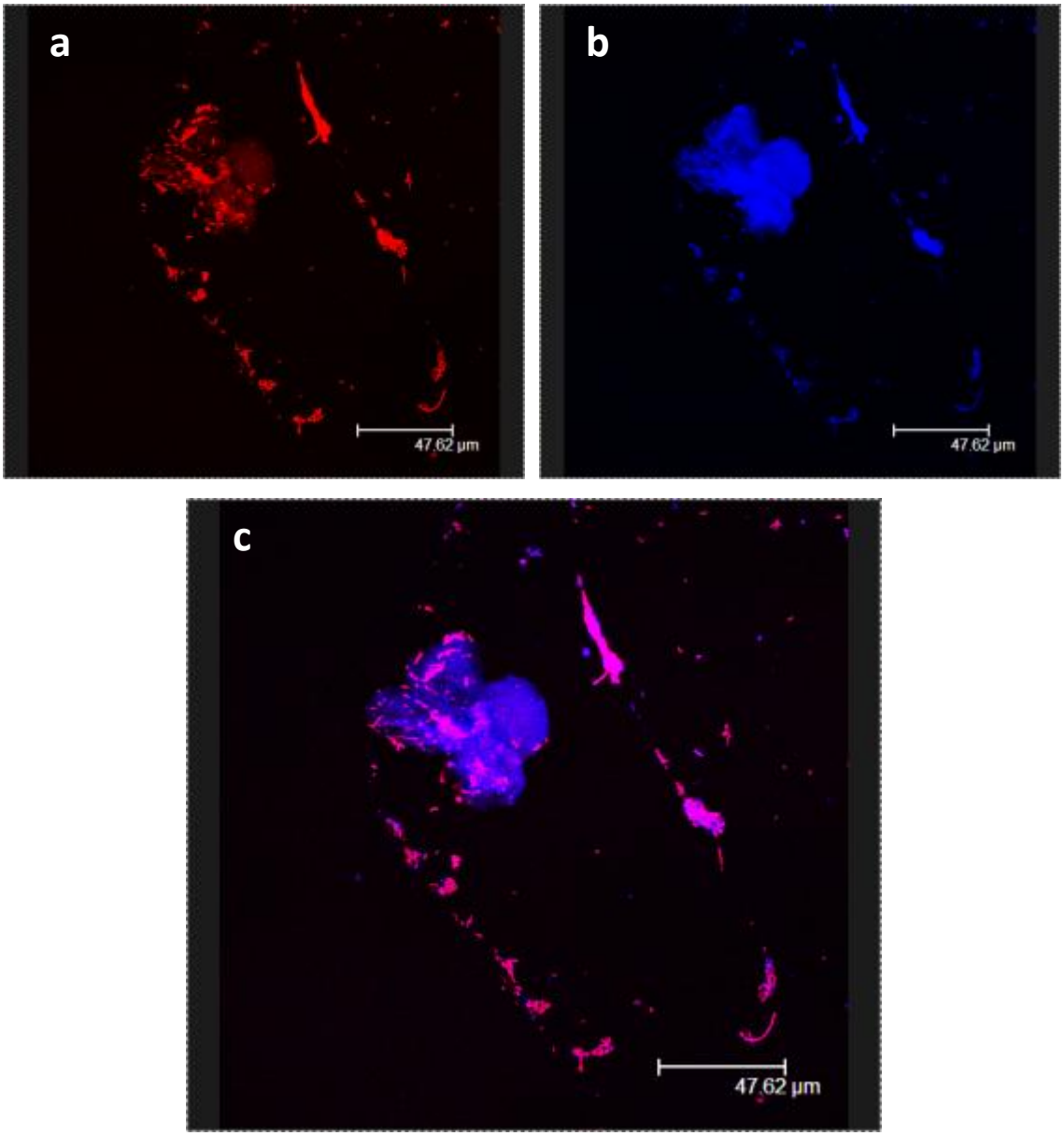


Figure 5.12: Fluorescence *in situ* hybridisation micrographs showing patient sample 025 with probes a) EUB338, b) Prev734, c) overlay image of both probe signals.

Microscopy results for patient sample 025 scoop 2 resulted in very few bacterial cells fluorescing (about 15 in the microscopic field) when using probe Bif698. Since the cells that were stained by the fluorochrome resembled cocci, it is possible that they were debris. Using probe Prev734 very few cells were found (about 7 in the microscopic field). This signal could represent cocci, debris or a mixture of both.

Microscopy on patient sample 025 scoop 3 did not result in any cells being found in any of the wells.

CLSM for patient sample 024 scoop 1 using probe Prev734 (Figure 5.13 a) resulted in detection of about 40 cells that could represent either cocci or short rods. Some debris or possibly small cell clusters were observed. A higher magnification confirmed cells to be short rods. Application of probe MUT590 (Figure 5.13 b) generally resulted in detection of few cells. Some cell clusters as well as relatively long chains of cocci were observed. Debris was in parts weakly fluorescing, but in other areas, strongly fluorescing. Application of universal probe EUB338 (Figure 5.13 c) revealed individual and clusters of cells spread over the microscope fields. Debris, sometimes a very bright signal, other times appearing as a cloud-like haze, was also observed. Cell morphology presented as cocci and chains of cells with a coccal morphology as well as rods and some filaments.

When both EUB338 and Prev734 were applied, microscopy focusing on EUB338 resulted in detection of many cells and cell clusters. Cells appeared to be mostly short rods, but some cocci, long rods and filaments or chains of cells were found to be fluorescing. As in the case of patient sample 025 signal for Prev734 appeared as staining debris. Individual cells that were stained using this probe appeared to be rods. Again the overlay of both probe signals showed more blue (Prevotella specific probe) than red (EUB338 universal probe) (image not shown).

Following these results laser settings were re-evaluated and it was found the laser exciting probe labelled with Cy3 also excited debris and emission of this was recorded by the laser set up for probe Cy5 in the multiplex FISH experiments, therefore resulting in a false-positive fluorescence.

Renewed microscopy of the multiplex FISH (Figure 5.14) following the adjustment of laser settings such as smart gain (brightness level of pixels) and smart offset (darkness level of the background) revealed clusters of cells of varying sizes as well as individual cells. The large cell clusters seemed to co-localise with debris, which appeared as a cloud-like haze. Morphology of the detected cells was mainly rods, with some cocci and some filaments. Focusing on probe Prev734 showed far less fluorescence than EUB338 and it appeared that most fluorescing structures were representing debris. A few cells looking like short rods or cocci could also be observed. The overlay image of both probes confirmed that the Prevotella probe nearly exclusively bound to debris, while some clusters near/on the debris showed individual cells fluorescing pink.

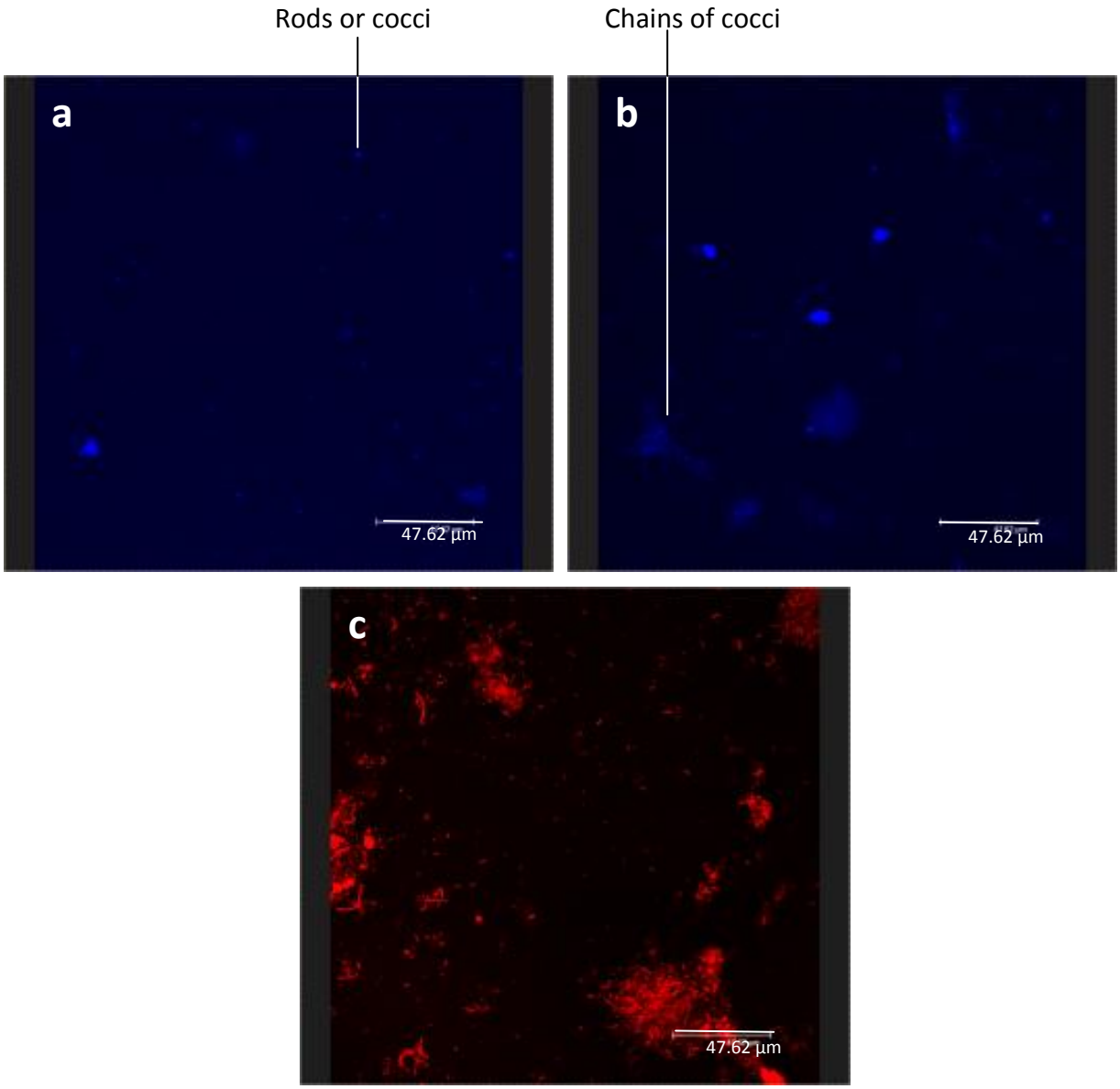


Figure 5.13: Fluorescence *in situ* hybridisation micrographs showing patient sample 024 with probes a) Prev734, b) MUT590 and c) universal probe EUB338

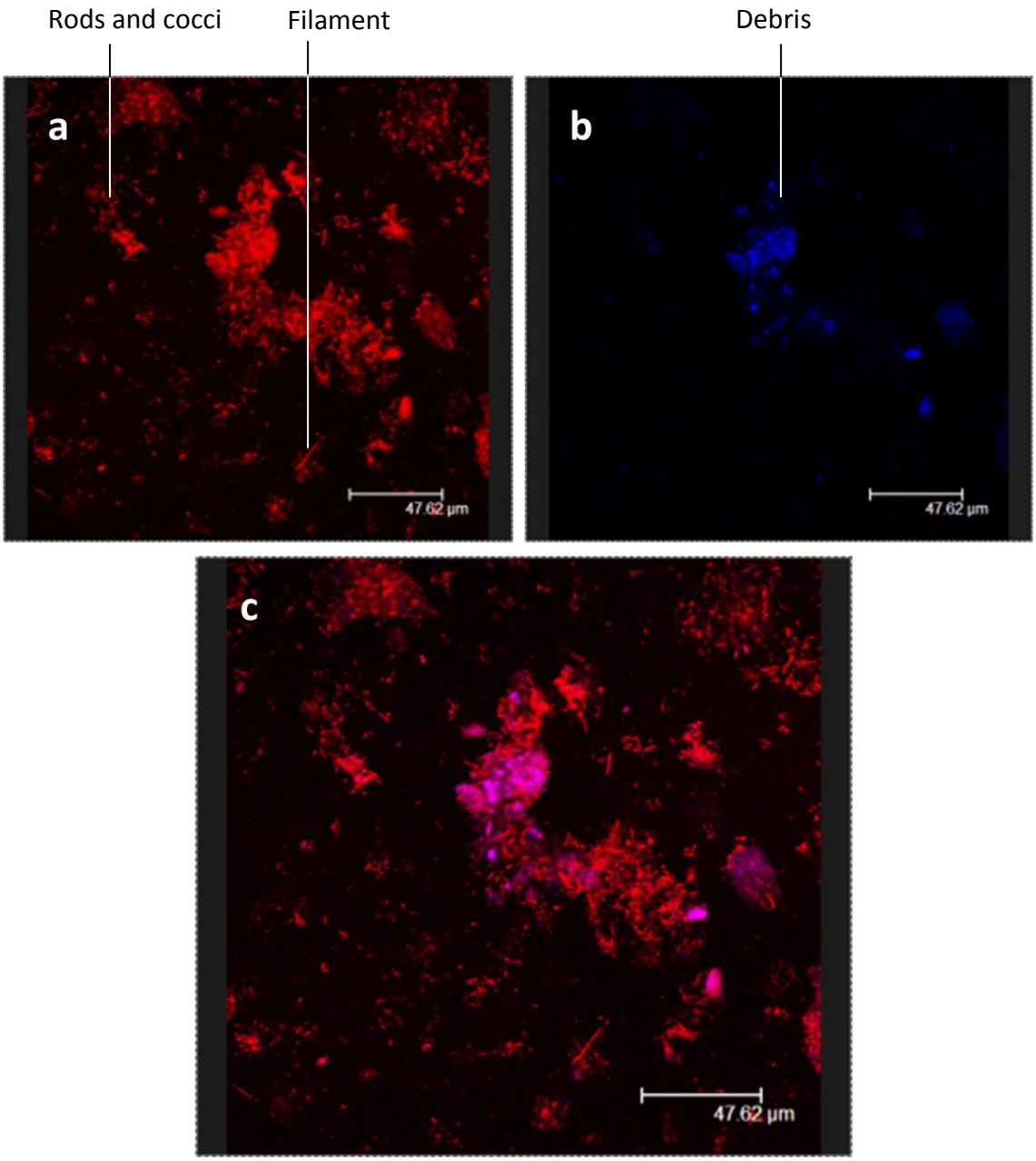


Figure 5.14: Fluorescence *in situ* hybridisation micrographs showing patient sample 024 with probes a) EUB338, b) Prev734, c) overlay image of both probe signals

Two further teeth were obtained (Figure 5.15 and Figure 5.16). Teeth were bisected and tooth halves photographed showing from where the sample scoops were taken. Microscopy of samples 026 and 027 is detailed in Table 5.11.

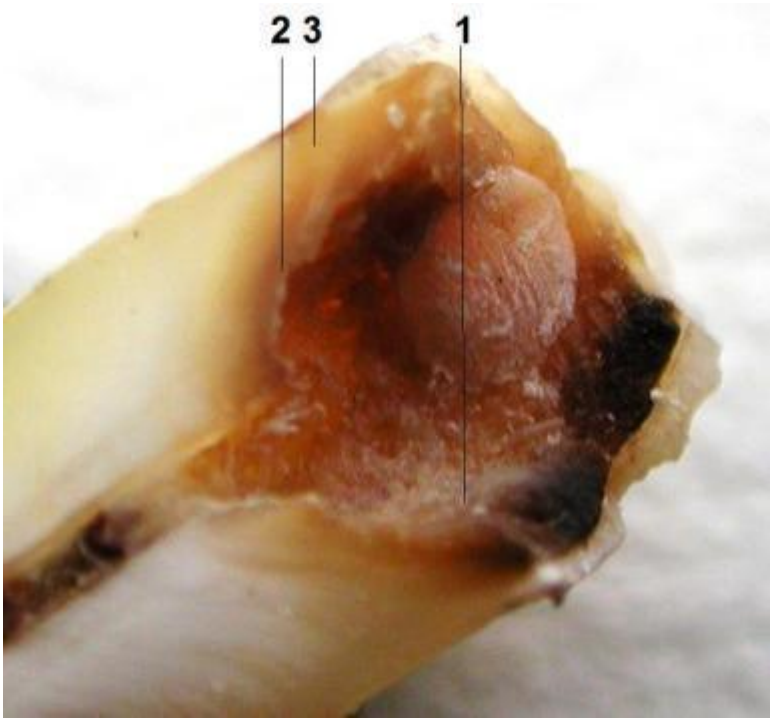


Figure 5.15: Tooth sample 026 with 1 indicating the sample from the innermost region of the lesion, 2 and 3 indicating sampling sites from the leathery and harder, progressing part of the caries lesion, respectively.

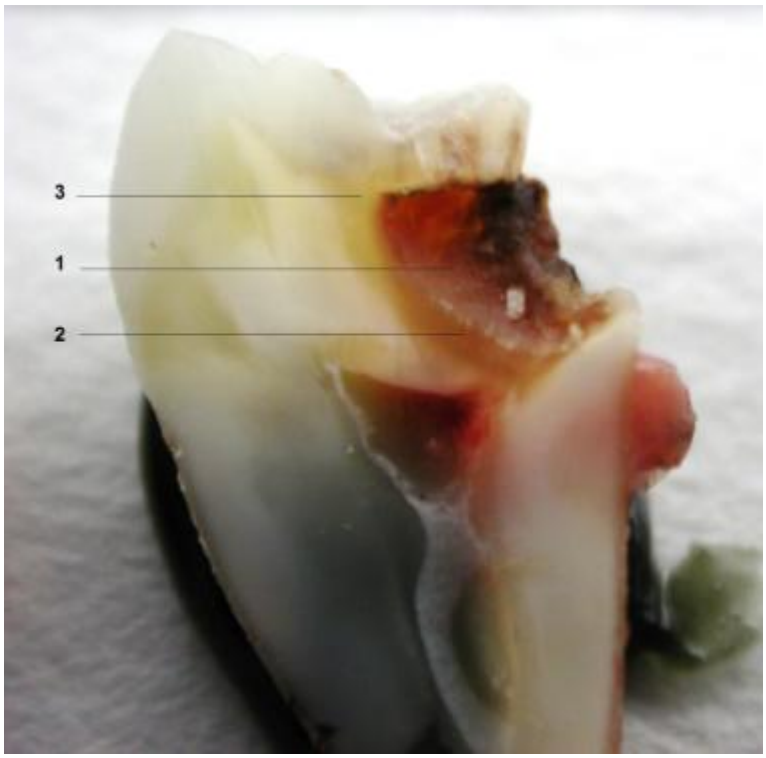


Figure 5.16: Tooth sample 027 with 1 indicating the sample from the innermost region of the lesion, 2 and 3 indicating sampling sites from the leathery and harder, progressing part of the caries lesion, respectively.

Table 5.11: Microscopy results of FISH on scoops taken from three areas of bisected carious lesions; +++ 300+ cells, ++ 100-300 cells, + 10—100 cells, +/- 1-10 cells, - 0 cells, BL bleedthrough, ND not determined

Sample ID		026 scoop 1	026 scoop 2	026 scoop 3	027 scoop 1	027 scoop 2	027 scoop 3
	Rods	+++	-	+	+++	+/-	-
EUB 338	Cocci	++	-	-	+	-	-
	Filaments	+/-	-	-	+	-	-
	Rods	-	-	+/-	++	+/-	ND
EUB 338 /	Cocci	-	-	+/-	+	-	ND
	Filaments	-	-	-	+/-	-	ND
Bif698		-	-	BL	BL	-	ND
	Rods	+++	-	-	++	-	ND
EUB 338 /	Cocci	++	+/-	-	+	-	ND
	Filaments	+/-	-	-	+/-	-	ND
Bif128		BL	-	-	-	-	ND
	Rods	++	-	-	+++	-	ND
EUB 338 /	Cocci	+	-	-	+	-	ND

Sample ID		026 scoop 1	026 scoop 2	026 scoop 3	027 scoop 1	027 scoop 2	027 scoop 3
	Filaments	+/-	-	-	+/-	-	ND
Prev282		BL	-	-	BL	-	ND
	Rods	+++	-	-	++	+/-	ND
EUB 338 /	Cocci	++	+	-	+/-	-	ND
	Filaments	+/-	-	-	+/-	-	ND
Prev734		-	BL	-	+	-	ND
	Rods	+++	-	+/-	++	-	ND
EUB 338 /	Cocci	++	+/-	+/-	+/-	-	ND
	Filaments	+/-	-	+/-	-	-	ND
MUT590		BL	-	BL	BL	-	ND
Bif128		++	ND	ND	ND	ND	ND
Prev282		-	ND	ND	ND	ND	ND
Prev734		-	ND	ND	ND	ND	ND
Mut590		+/-	ND	ND	ND	ND	ND

Microscopy of sample 026 scoop 1 EUB338 showed numerous bacteria, many rod shaped and some were long rods or filaments. Cells were spread relatively evenly over the microscope field and no clusters were seen, but some debris.

The second well representing hybridisation using probes Bif698 and EUB338 did not result in the detection of any cells with either probe. Multiplex FISH using Bif128 and EUB338 resulted in detection of cells and some debris for the universal probe, but no fluorescence representing cells could be seen with the bifidobacteria probe. The combination of Prev282 and EUB338 showed rods and cocci (~ 40 – 60 cells in the microscopic field), but the only fluorescence detected using the Prevotella probe turned out to be bleed-through from the other laser. Similar results were found with Prev734 and EUB338, where plenty of individual cells – rods of different lengths, cocci, as well as some debris was seen using the universal probe, but no cells could be detected using the specific probe. Only the combination of MUT590 and EUB338 resulted in detection of cells for the universal probe and possibly fluorescence from a cluster *S. mutans*. It was not possible to exclude the possibility that this signal was unspecific probe binding to debris.

Application of probes Bif128, Prev282, Prev734 and MUT590, resulted in detection of possibly short rods for Bif128, no cells for Prev282 or Prev734 and ~ 9 fluorescing points on the image for MUT590, which could represent cells rather than debris.

No cells were detected for scoop 2 using probe EUB338 on its own, Bif698 + EUB338 and Prev780 + EUB338. Combination of Bif128 and EUB338 resulted in a few specks fluorescing on the universal probe image that could represent either short rods or debris. Nothing was detected with the bifidobacteria specific probe (Bif128). Use of Prev734 and EUB338 showed about 30 cells that could be rods and short rods or cocci for the universal probe and nearly the same signals for the specific probe indicating that the signal represents unspecific binding of both probes to debris. The same fluorescence was observed for probe combination MUT590 and EUB338.

Individual probes applied to this sample slide were not examined.

Scoop 3 of patient sample 026 resulted in very weak signal of about 10 – 15 cells for the EUB338 probe alone. All other multiplex combinations resulted in either no signal

at all (Bif128+EUB338, Prev282+EUB338 and Prev734+EUB338) or detection of very few cells for the universal probe and no signal or bleed through for the specific probes (Bif698+EUB338 and MUT590+EUB338).

CLSM of patient sample 027 scoop one using probe EUB338 showed many individual cells spread relatively evenly over the microscope field and only a few clusters were observed. Morphology of the detected cells was made up of approximately 90 % rods. Microscopy of Bif128, Bif698, Prev282 and MUT590 each paired with EUB338 resulted in detection of mostly rods, sometimes in clusters when focusing on the universal probe, but the specific probes resulted in either no signal at all or it was shown that signal was bleed-through from another laser.

The only exception to this was the probe combination of Prev734 and EUB338. Using the universal probe, short and long rods were detected, including one small cluster of long rods (Figure 5.17 a). Signal from the *Prevotella* specific probe resulted in a signal similar to the cluster of rods seen with the universal probe (Figure 5.17 b). The overlay image confirmed this as seen by the pink staining in the micrograph. At higher magnification, the cluster seemed to be made up of about 10 cells surrounded by another seven individual cells. The image for probe Prev734 shows that only the larger rods in the cluster fluoresce with this probe and none of the individual cells surrounding the cluster emitted a signal. Tests to exclude the possibility of bleed-through in this case proved that the detected signal using the universal probe EUB338 paired with the *Prevotella* specific probe Prev734 resulted in true signal.

Since microscopy of patient sample 027 scoop 2 resulted in detection of a maximum of 1 or 2 cells or debris per microscopic field for probe/probe combinations 1 - 6, wells 7-10 were not examined. Microscopy of patient sample 027 scoop 3 did not result in any cells being detected using the universal probe. Consequently, wells 2 – 10 were not examined.

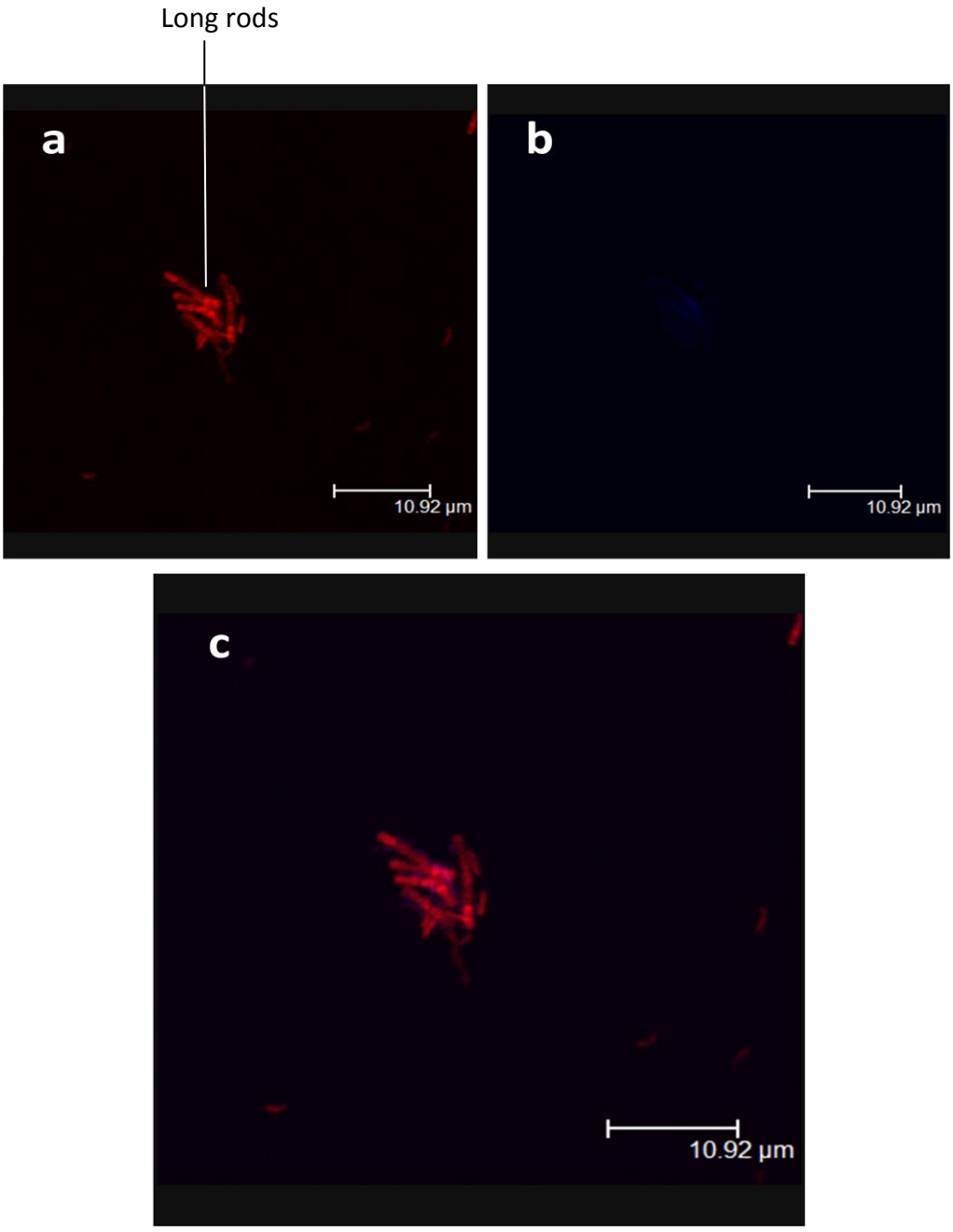


Figure 5.17: Fluorescence *in situ* hybridisation micrographs showing patient sample 027 scoop 1 with probes a) EUB338, b) Prev734, c) overlay image of both probe signals.

Tooth number 028 was acquired, bisected and photographed (Figure 5.18).

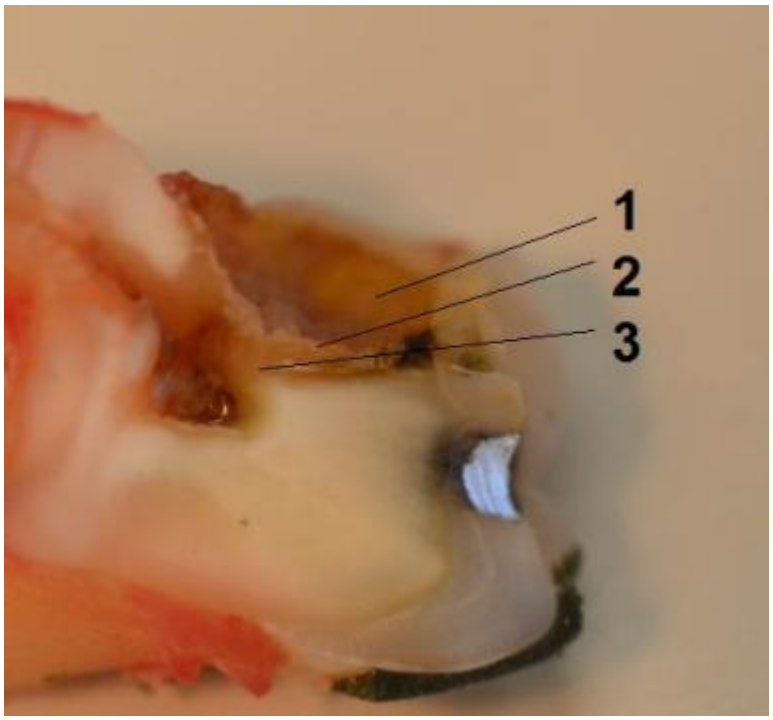


Figure 5.18: Tooth sample 028 with 1 indicating the sample from the innermost region of the lesion, 2 and 3 indicating sampling sites from the leathery and harder, progressing part of the caries lesion, respectively

Microscopy of the Gram-stained scoop samples revealed many bacterial cells in all three sites from which the scoops were taken (Figure 5.19 1-3). The majority in all samples seem to be rods, but some filaments and chains of cocci were observed. Gram-positive as well as Gram-negative staining bacteria were present.

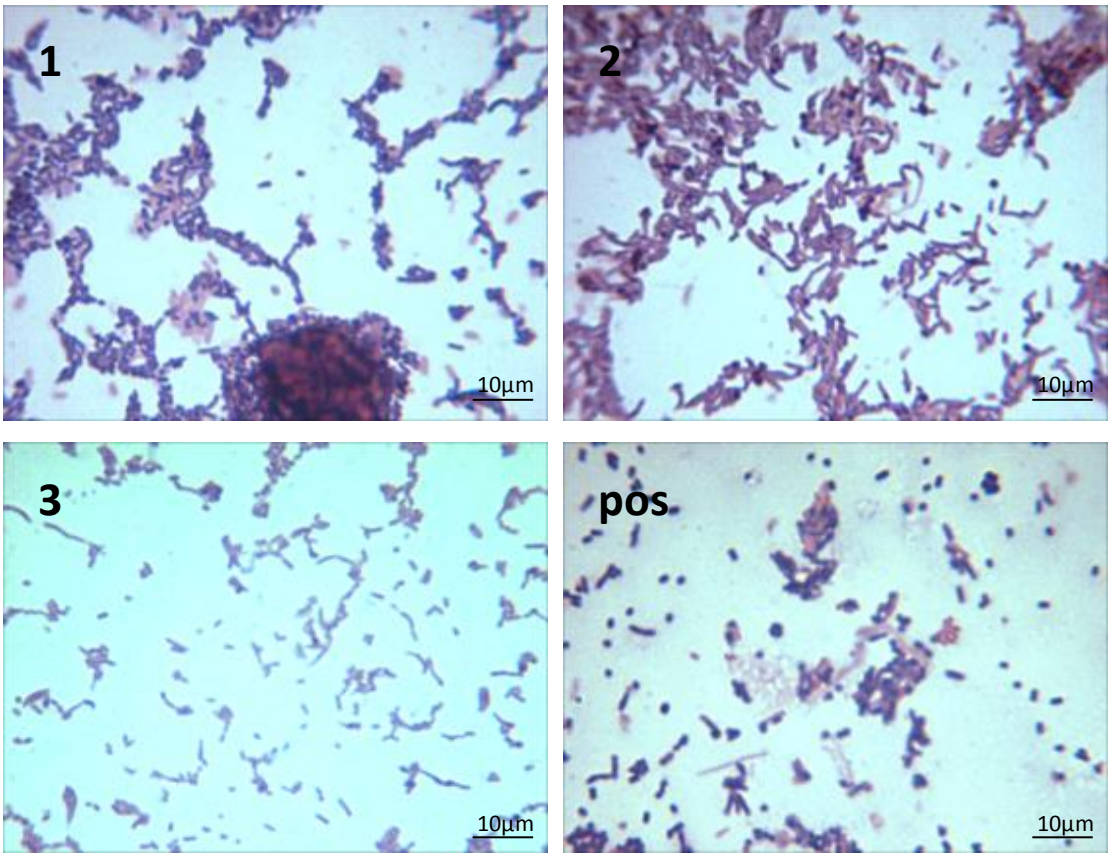


Figure 5.19: Gram stained scoops 1, 2 and 3 and the plaque positive control (pos)

CLSM of the remaining tooth hybridised with the universal probe EUB338 showed fluorescence that resembled bacterial cells – mostly rods, and slight background staining of the dentine (Figure 5.20). The surface appeared as if a smear was wiped over the surface (Figure 5.20 a and b), which was most likely artificial residue from the saw blade when sectioning the tooth. Higher magnification showed longer rods in the top half of the image, while in the lower half, fluorescence possibly represented shorter rods, cocci or unspecific hybridisation of the probe with dentine. The bright fluorescing area in the centre of the image possibly represented a cell cluster of debris on the surface of the bisected plane of the tooth.

Inspection of another microscope field did not show a smear-like effect like the first, but this field was the exception amongst all examined fields, which all showed this smear-like effect to some degree. Individual cells appeared evenly spread, with the exception of one big cluster. Further magnification of that cluster showed the cluster to consist of rod-like structures mixed with brighter fluorescing areas.

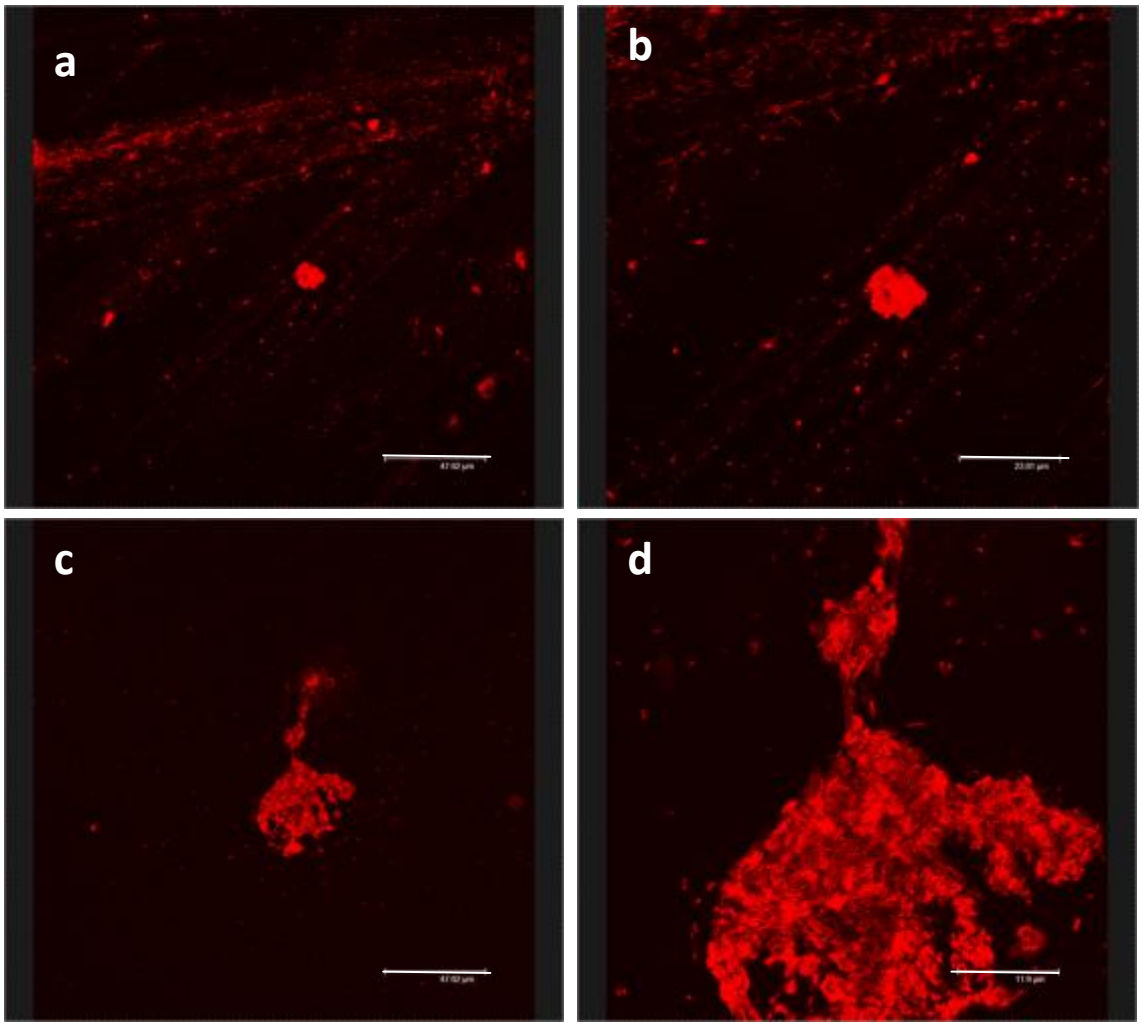


Figure 5.20: Fluorescence *in situ* hybridisation micrographs showing patient sample 028 with probe EUB338 used in all images. b) and d) are magnifications of a) and c), respectively.

5.5 Discussion

Fluorescence *in situ* hybridisation probes of the most prevalent genera (*Prevotella* and *Lactobacillus*) and one high G+C genus (*Bifidobacteria*) were designed and validated. A species-specific probe from another study (Thurnheer et al. 2001) as well as the widely used universal eubacteria probe EUB338, were included. FISH probes targeting *Prevotella*, *Bifidobacteriaceae*, *S. mutans* and EUB338 were applied to tooth samples. Application of probes Bif128, Bif698, Prev734 and MUT590 resulted in signal, which might have represented cells. In multiplex experiments only, the combination of the universal probe EUB338 together with Prev734 resulted in true signal, other positive signals turned out to be bleed-through of one laser resulting in excitation of the fluorochrome by the wrong laser. Although application of probes resulted in fluorescence of target organisms in the patient samples only in a few instances, results indicated probes bound to the correct target and issues that needed addressing related to sample-taking and user experience when performing microscopy rather than the probes having failed.

FISH on bacterial smears using EUB338 was successful, but FISH on bisected teeth using EUB338 did not result in detection of bacterial cells. This suggested that somewhere during fixation and permeabilisation of cells and hybridisation, the protocol failed. Microscopy may have been unsuccessful due to the complexity of having a tooth rather than a well of a microscope slide.

As an intermediate step between bacterial smears and microscopy of bisected teeth, FISH on hydroxyapatite discs using an adapted protocol from Thurnheer et al. (2004) was attempted. The protocol used in this study used shorter incubation times (6 - 16 h) than the 16.5 - 64.5 h in the other study, but since Thurnheer et al. (2004) detected cells at 16.5 h it was expected to successfully detect cells after 16 h, if not at 6 h. Discs in this study were not pre-conditioned with saliva and the nutrient broth used differed slightly. Differences in cell fixation methods may furthermore be responsible for the differing results of the experiments. However, since this experiment was intended as a stepping stone between bacterial smears on slides and microscopy of teeth and it was not a focal point of this study to develop an *in vitro* model, efforts of growing cells on HA discs were soon abandoned and other avenues sought to address the remaining issues.

Taking scoops from the caries lesions and processing them in a similar way to the scoops that were taken to create the clone and isolate libraries proved successful. Applying only EUB338 to the scoops taken from the three patient samples revealed large numbers of bacteria for two of the three samples. Since the samples showing the least and the most cells on the micrographs were taken within minutes of each other, processed immediately and the same way, handling is unlikely to have caused this result. Another possible explanation could be the different states of caries progression for each patient and the bacterial load, which varies for each individual as described in a study on the development of plaque over two months in six subjects where it was found that some volunteers were slow/poor plaque formers throughout the study period (Listgarten et al. 1975). The finding that two of the patient samples showed opposite proportions of rods and cocci, while the sample with the least detected cells showed a mixture of rods and cocci, but no filaments, which were seen in limited numbers with both the other samples, support this theory.

Intra- as well as inter-individual variability of bacterial adhesion and surface colonization has been observed in several studies (Diaz et al. 2006; Dige et al. 2007; Al-Ahmad et al. 2009) and most likely had the biggest influence on the findings in this study. This point is highlighted especially by the study by Diaz et al. (2006) which presented confocal micrographs of two patients with differences in proportional biomass comparable to those observed in this study. Al-Ahmad et al. (2009) concluded from this that not only biofilm formation but also initial adhesion of bacterial cells occurs in a subject-dependent matter. It would be a logical conclusion that the progression of a caries lesion, too, was influenced by each individual's microbiota, which in turn is influenced by genetics as well as environmental factors, such as diet.

As mentioned in the results section, observed differences in fluorescence intensity represented the availability of RNA to which the FISH probes bound. Odaa et al. (2000) observed in their study examining the influence of growth rate and starvation on FISH of *Rhodopseudomonas palustris*, that during starvation there was a correlation between RNA content and percentage of hybridised cells. They concluded that 16S rRNA-targeted probing should be used with caution for quantitative detection of populations since cells are influenced by their physiological history as well as their current physiological state. This, however, more likely applies to nutrient-restricted

habitats such as marine, freshwater or soil and is probably of limited consequence to the relatively nutrient-rich oral cavity.

Design of the specific probes was successful and between four and five probe candidates were designed for each genus/family. It is known that achieving good fluorescence in FISH experiments on lactobacilli is difficult due to their cell wall's resistance to probe penetration (Quevedo et al. 2011). Indeed, validation of probe Lacto379 resulted in weak fluorescence signal of the narrow panel and positive control experiments using EUB338 on *L. gasseri* and *L. vaginalis* resulted in fluorescence of about 40 - 60 % of the cells on the slide. Validation of probes Bif698 and Bif128 showed good fluorescence so that probe Bif1452 was not validated.

All *Prevotella* probes were validated. Probe Prev1528 was discounted since it did not appear to hybridise to *P. tannerae.* Probes Prev512 and Prev781 resulted in weak fluorescence of some control species. During the validation process for probe Prev734 few individual cells of *B. zoogleoformans* and *T. forsythia* were lighting up non-specifically, but numbers were so few, especially compared to the strong fluorescence of the target organisms that fluorescence of probes Prev282 and Prev734 was deemed strong and specific enough to be used on teeth.

Validation of probe MUT590, designed by Thurnheer et al. (2001), resulted in fluorescence of *F. nucleatum*. This was unexpected, since this species was also included in the other study's validation panel and was recorded as negative. Reasons for this are not clear. It is questionable that use of a different subspecies would be able to account for this result. The only major difference in the protocol was the method of permeabilisation of cells, which was achieved by lysozyme treatment in the study by Thurnheer et al. (2001), compared to the 1:1 PBS/EtOH treatment used in this study. Since cell morphology of *S. mutans* and *F. nucleatum* are different, it was decided to not investigate this issue any further, but to include this probe in experiments as planned.

The first tooth samples resulted in successful detection of bacteria using all probes in the case of sample 025 and all except Bif698 for sample 024. Since the targeted genera/species make up only a fraction of the total bacterial load, results showing far fewer cells fluorescing using the specific probes compared to the universal probe

tallied with expectations. Therefore, it was surprising to see in the multiplex FISH of EUB338 and Prev734 that seemingly more fluorescence originated from the specific probe. It appeared that Prev734 bound preferentially to debris, but it became clear that laser settings had to be adjusted to avoid bleed-through from the other laser. Following revision of settings, multiplex FISH indeed resulted in far more signal stemming from the universal probe than the specific probe. Nevertheless, it still appeared that Prev734 preferentially bound to debris and fluorescence did not, or not exclusively, represent bacterial cells.

It was expected to find more bacteria in the heavily infected dentine compared to the scoops taken further toward and directly from, the affected dentine, but it was unexpected to find few or no cells at all in scoops two and three. It was not clear if only few bacteria were in the advancing front and discolouration of the dentine was due to chemical reactions. Although there is currently no definite answer as to the extent to which bacteria progress in the advancing lesion (Kidd et al. 1993; Banerjee et al. 2002), studies looking at the deep layers of advanced caries lesions did detect various phylotypes using culture as well as molecular techniques (Byun et al. 2004; Munson et al. 2004; Nadkarni et al. 2004; Aas et al. 2008; Lima et al. 2011) and detection of bacteria with, at the very least, the universal FISH probe at all levels of cavitation was expected.

Munson et al. (2004) reported carious dentine microbiota to be dominated by Gram-positive taxa, such as streptococci and that *S. mutans* was detected in 60 % of patient samples. Moreover, Aas et al. (2008) reported the predominance of specifically *S. mutans* in deep dentine caries lesions, making the detection of the species in this study using the *S. mutans* specific probe relatively likely. On the other hand, Aas et al. (2008) suggested that *S. mutans* seems to have a more dominating role in caries lesions of primary dentition, compared to secondary dentition and in a study by Lima et al. (2011) only 44 % of the 22 streptococci-positive patients (out of 27 patients in total) were positive for *S. mutans* with predominance levels of 11 % in the deep layer of the lesion. It could therefore be that the limited number of teeth used in this study simply contained few *S. mutans,* which seemingly were only located in the superficial layer of the lesion.

Yang et al. (2012) noted that an overpopulation of *Prevotella* in saliva distinguished caries microbiota from healthy ones and that caries- and health-associated populations were made up of different arrays of *Prevotella* species. However, one cannot necessarily conclude from presence in saliva of caries-active patients, that these species can also be found in the deep dentine lesion. Nonetheless, detection of *Prevotella* and *Prevotella*-like species have been reported in advanced caries lesions (Martin et al. 2002; Nadkarni et al. 2004; Chhour et al. 2005). Martin et al. (2002) examined carious dentine of lesions in which the lesion was so deep that the pulp was involved and reported a high incidence of *Prevotella*-like bacteria. In a follow-up study Nadkarni et al. (2004) targeted *Prevotella*-like taxa in a similar sample pool with genus-specific primers and found numerous phylotypes that grouped into clusters, all of which were represented in most carious-dentine samples.

Results of these studies led to the expectation that using a specific probe, as was accomplished here, would enable detection of this genus in advanced caries lesions. It is unlikely that the probe did not bind, since probes resulted in fluorescence signal in some of the samples and *Prevotella* have been visualised in other studies (Diaz et al. 2006).

Detection of *Bifidobacteria* was also thought possible, since studies, especially those targeting *Bifidobacteria* specifically, reported a high detection rate for this genus (Modesto et al. 2006; Mantzourani et al. 2009; Kanasi et al. 2010; Tanner et al. 2011). In addition, Beighton et al. (2010) reported isolation of *Bifidobacteria* from caries lesions in 96.8 % of subjects, while detection of MS, lactobacilli and yeasts in the same subject pool were lower at 91.7 %, 90.4 % and 60.3 %, respectively. What's more, Beighton et al. (2010) argued that since *Bifidobacteria* are found mainly in active lesions (Becker et al. 2002; Mantzourani et al. 2009) they might be primarily involved in disease progression, which means that detection of this family in advanced caries lesions, as studied here, has a high potential.

Once probes have been validated and the protocol adapted, if necessary, the remaining target for FISH probes, the lactobacilli, should be detectable equally in deep dentine lesions due to their association with caries, and especially deep dentine lesions. In many studies, including this one, lactobacilli are part of the predominant

species and often a high variety of taxa can be found within the same lesion (Munson et al. 2004; Chhour et al. 2005; Lima et al. 2011).

One problem encountered in this study was the definite and potential binding of specific probes to debris. To try to reduce unspecific binding of probe to debris, to reduce the amount of debris overall and possibly thereby release bacteria from the debris, sample 026 was subjected to manual crushing of the debris, while both samples 026 and 027 were vortex mixed for 5 minutes. Slides of both samples underwent an extra step in the FISH protocol whereby hybridisation buffer was applied to wells in order to block sites to which the probe could non-specifically bind.

Despite these measures, microscopy revealed similar findings to the previous samples. Scoop 1 for both samples resulted in detection of cells in most cases where probe EUB338 was applied alone or in conjunction with one of the specific probes. However, microscopy of the specific probes either resulted in no detection of any cells or detection of fluorescence that was likely to be unspecific binding to debris or bleed-through from another laser. Overall, there seemed to be a reduction of fluorescence that looked like debris, especially relatively big chunks compared to the previous samples. However, the different treatments of samples 026 and 027 of manually crushing and vortex mixing or vortex mixing alone, did not result in any discernible effect when the two samples were compared. The newly introduced step of blocking unspecific binding of probe seemed to have had some effect, since less cloud-like debris or background fluorescence was observed.

Amongst all the reactions not resulting in detection of the targeted genera and species in the multiplex FISH experiments, there was one very promising outcome, nevertheless. The multiplex reaction of probes EUB338 and Prev734 applied to sample 027, scoop 1 resulted in seemingly true fluorescence of about five cells for the specific probe amongst the approximately 130 cells fluorescing from probe EUB338. This would suggest fluorescence of the specific probe represented a little less than 4 % of the total number of bacteria fluorescing. Furthermore, it was established that this was not a result of bleed-through. Microscopy of scoops 2 and 3 for both samples again resulted in very little or no emitted signal.

There has been much debate as to whether / how far bacteria progress in the advancing front of the caries lesion. Banerjee et al. (2002) used FISH to detect and enumerate bacteria in human carious dentine at four different levels of caries lesion progression. At the various layers of the lesion bacterial counts ranged from 7.34 x 10^6 cfu/mg dentine at the superficial layer to 1.69 x 10^6 cfu/mg dentine at the excavation front,. The final sample taken from each patient, labelled the advancing front, was beyond the point where a dentist would normally terminate excavation and 3.4 x 10^5 cfu/mg dentine were recorded. Quantities of bacteria in that study were much higher than those attempting enumeration of cells previously, but significant differences between studies existed. van Strijp et al. (1994), for example, recorded 0.11 – 29 x 10^5 cfu, although the dentine was not weighed and the dentine was furthermore decalcified before slabs were carried by study participants *in situ* for seven weeks. Kidd et al. (1993) recorded ≤ 100 cells in nearly 40 % of the samples taken from the superficial layer of a caries lesion. However, both studies used culture techniques and since less than 50 % of oral species can be readily grown in the laboratory, fewer numbers would be expected. Also, Kidd et al. (1993) used a burr to take samples, thereby possibly damaging bacteria during the collection process. Furthermore, it appears that culture medium was not pre-reduced and sitting on ice while samples were taken resulting in anaerobic bacteria being subjected to oxygen concentrations that could be harmful.

A possible explanation for the high cell numbers in the study by Banerjee et al. (2002) could be contamination of lower excavation levels from those higher up, since samples were taken from the heart of the lesion subjacent to the enamel-dentine junction, working down to the advancing front. Investigators argued that this effect was reduced by using fresh hand excavators and the scraping technique used, but it is unlikely that this would have avoided the issue to the same degree as bisecting the teeth and taking samples from the cut plane of the tooth would have.

To address several issues, such as experience of the investigator regarding sample procurement and if bacteria are or are not present in areas further into the dentine than just the superficial layer, a further tooth was acquired and bisected. In this instance, a trained dentist took the three sample scoops. Instead of FISH, scoops were Gram stained following vortex mixing, and microscopy revealed bacteria of various

morphotypes staining Gram positive and Gram negative from all three levels of lesion progression. Gram staining relies on the density and morphology of the cell wall and live cells cannot be distinguished from dead ones, but if bacteria can be found in the advancing front it would be plausible that the majority of these cells are alive, since they are driving the disease progression. These results suggest that more cells should be detected using FISH and CLSM and that tooth preparation and following FISH methodology need further development.

Since FISH on scoops taken from caries lesions was relatively successful in principle, the remaining tooth half that was not used for the Gram stain was used for FISH with only probe EUB338 applied directly to the tooth surface. In this last experiment, visualisation of cells on the tooth surface was indeed successful and it was possible to distinguish individual cells and cell clusters. One shortcoming of this set up was that it was not possible to say to which area of the lesion and therefore, which level of discolouration of dentine the fluorescence related. Nevertheless, the successful application of a FISH probe directly to the caries lesion of an extracted tooth is a very promising step for future studies in trying to avoid cumbersome and, by default, limited *in vitro* or *in vivo* studies.

It is not known whether any other study has tried to visualise bacteria in the progressing caries lesion directly on a bisected tooth. Other studies have presented successful FISH experiments on embedded pulp sections in which dentine tubules were visible (Nadkarni et al. 2010; Rechenberg et al. 2011), but embedding of the relatively hard dentine of the advancing front is not possible and experiments are therefore not comparable. Some studies have used excavated sample solutions spotted on microscope slides as presented in this study (Banerjee et al. 2002) whilst others have made use of glass (Dige et al. 2007), bovine enamel (Al-Ahmad et al. 2009) or bovine dentine slabs (Jung et al. 2010) worn by volunteers *in situ* for set periods of time. However, the latter studies used those methods to characterise early colonisation, and not advanced caries lesions and how bacteria (co-)localise and possibly interact in those lesions, as was the aim of the current study.

Many studies have used 4 % paraformaldehyde, a cross-linking agent, for fixation of cells and enzymatic treatments, such as lysozyme, to facilitate permeabilisation of the

cell wall (Amann et al. 1995; Thurnheer et al. 2004; Diaz et al. 2006). However, it was found that paraformaldehyde treatment can be detrimental for whole-cell hybridisation of Gram positive bacteria (Amann et al. 1995) and some studies used a staggered combination of paraformaldehyde and ethanol fixation (Al-Ahmad et al. 2007), while others used ethanol fixation only (Thurnheer et al. 2001). It was found that for cell fixation and permeabilisation using an ethanol/PBS solution (50 % v/v) worked equally well for Gram positives as well as Gram-negative species and was the method of choice in this study. The disadvantage of using the combination of paraformaldehyde and lysozyme treatment could mean that Gram-positive bacteria are overlooked in the analysis and lysozyme, which only works on sensitive cell walls, may not actually allow permeabilisation of all cells. Good results were achieved with the presented protocol, but in the case of *Lactobacillus* it may become necessary to incorporate a lysozyme treatment in the protocol. In future studies, it might be a possibility to subject one-half of the tooth to the original protocol and develop another for species that require enzymatic treatment in order to permit improved probe uptake and therefore fluorescence.

5.6 Conclusion

It can be concluded from the results obtained in this study that microscopic examination of the microbiota in carious lesions using *in situ* hybridisation is challenging. The design of primers appeared relatively straight forward, since particular specifications had to be adhered to but repeating experiments, such as FISH on HA discs, proved more difficult than anticipated. As a biological sample, caries lesions naturally vary in their dimensions, colour and texture. It was therefore difficult to assign uniformly the excavation points with regards to observed level of staining/hardness of tooth matrix and some variation in data had to be expected. Other limitations were presented by simple physics in that peaks of probe spectra had to be far enough from each other to avoid bleed-through, as observed in this study, and more than three dyes cannot be used at any one time.

Despite these restrictions, newly designed oligonucleotide probes for one family and one genus were applied successfully to material excavated from carious lesions resulting in fluorescence, showing that probes are specific and that metabolically active bacteria are present in the lesion. Multiplex FISH experiments require further

development, but results from joint application of probes EUB338 and Prev734 show that specific genera could be detected. The final experiment showed that fluorescence *in-situ* hybridisation directly on bisected teeth is possible and further experiments fine-tuning the methods used in this study are required.

Chapter 6

General discussion

6 Chapter 6

Despite intensive research and the availability of many oral care products, dental caries remains a substantial global health problem, with an associated economic burden. The majority of microbiological studies into caries have focused on *S. mutans* (Clarke 1924; Hamada et al. 1980). Recent advances in culture-independent analysis have shown that the caries-associated microbiota is far more complex than previously thought, and there clearly is a need to understand the interactions between bacteria in the progressing lesion at the molecular level in order to recognise the key mechanisms causing lesion initiation and progression (van Ruyven et al. 2000). The so-called "low pH" streptococci as well as *Actinomyces* and *Bifidobacterium*, for example, are thought to contribute to the demineralisation of the tooth matrix due to their high acidogenic potential (van Houte et al. 1996; van Ruyven et al. 2000). However, biases in 16S rRNA-based methods have been identified which have resulted in genera, such as *Actinomyces* and *Bifidobacterium* belonging to the high G+C phylum *Actinobacteria,* being underestimated in molecular studies (Munson et al. 2002; Munson et al. 2004; de Lillo et al. 2006).

Consequently, this study examined comprehensively the oral microbiota found in carious lesions employing molecular methods that aimed to minimise bias against this phylum by using newly-designed primers on the one hand and techniques that inherently avoid biases, such as cloning, on the other. Furthermore, a protocol was developed for the visual assessment of prevalent caries species within the layers of the lesion to elucidate (co-)localisation, which ultimately might reveal interactions amongst species that are responsible for disease progression.

Both the established Sanger sequencing-based method as well as the recently-developed pyrosequencing approach revealed a highly diverse microbiota of similar range with as-yet undetected species present. However, neither molecular technique improved detection levels of *Actinobacteria* to those seen with culture. Moreover, it was evident that both techniques have associated limitations, which affect findings in different ways. Sanger sequencing, for example, relies on cloning that has the potential weakness of preferential ligation (Taylor et al. 2007), whilst at the same time being labour intensive and allowing analysis of only relatively few samples. Pyrosequencing on the other hand, can be affected negatively by homopolymers that can lead to

erroneous base calling as well as the formation of chimeric sequences. Both can lead to the overestimation of species richness (Keijser et al. 2008). The development of a novel FISH protocol for the visual examination of bacterial species directly on teeth was successful in principle, though additional developmental work, such as improvement of the methodology with regards to probe penetration and efficient hybridisation, is required.

Improved detection of all species resulting in a true reflection of the microbiota present in dentine lesions holds the potential to challenge current views of disease initiation and/or progression. It is generally accepted that not a single species is responsible for dental caries (Kleinberg 2002; Takahashi et al. 2008), but that any species capable of lowering the pH beneath the critical threshold of 5.5 can contribute towards disease initiation (Kleinberg 1961). Therefore, current research aims to assess not just the presence of species, but also phenotypic characteristics, such as the aforementioned acidogenicity and acidurance of organisms (van Ruyven et al. 2000; Takahashi et al. 2011).

The investigation into the source of bias influencing preferential amplification of low G+C sequences when both low and high G+C sequences are present revealed that neither the DNA extraction method nor the choice of polymerase represented the sole cause for the biased amplification. Existing primers were adapted to reduce bias that can occur due to mismatches (Sipos et al. 2007), whilst additional novel primers were also designed. During the primer validation process it was confirmed that different primer pairs exert biases (Frank et al. 2008), though primer bias was expressed differently in conjunction with the two polymerases. Amplification of mock communities consisting of low and high G+C species using the same primer pairs resulted in the overestimation of the high G+C organism using *Taq* polymerase, while use of the Phusion polymerase resulted in underestimation of the G+C-rich species. The forward primer in this example had been shortened at the 3’ end to avoid a mismatch with the low G+C organism. However, it has been shown that intrinsic sequence characteristics can influence amplification (Hansen et al. 1998), which may be the case here, resulting in a bias against the low G+C organism. The Phusion polymerase does not seem to be as sensitive to factors such as mismatches (see discussion Chapter 2), which could be due to the way the molecule is constructed or

additives that are included in the supplied high G+C buffer that was used. In conclusion, results clearly showed that the crucial factor in detection of species from mixed communities was the choice of primers and not the *Taq* polymerase as had been suggested previously (Varadaraj et al. 1994).

Findings from the molecular analyses of patient samples, consisting of excavated carious dentine, confirmed the hypothesis that different primer combinations result in detection of different species for each primer pair (Baker et al. 2003; de Lillo et al. 2006), while the bias against the high G+C phylum *Actinobacteria* was still evident with every primer combination when compared to culture. Some of the novel sequences were detected with several different primer combinations, others with only one, illustrating that the use of fewer primer combinations would have resulted in detection of fewer novel sequences. Furthermore, the importance of the combination of culture and molecular analyses was illustrated by the discovery that 66 % of taxa could only be found using molecular methods (including all novel sequences), while a little over 5 % of taxa were exclusively detected using culture. Indeed, results of this study confirm findings of other recent studies in that culture analysis complements molecular analysis (Kanasi et al. 2010; Tanner et al. 2011), implying that culture analysis should be included in future studies. Overall, the data supports the hypothesis that one primer pair is not sufficient for the exhaustive characterisation of the oral microbiota. The findings of a systematic *in-silico* review of bacterial 16S primer specificity support this suggestion, since bacterial sequences deposited in databases at the time were found to vary substantially, making the design of a sufficiently long primer matching all bacterial 16S rRNA gene sequences impossible (Baker et al. 2003).

The analysis of the six patient samples with the same primer combinations, but the Phusion polymerase instead of the *Taq* polymerase, was beyond the scope of this study. It would, however, be very interesting to determine if use of the Phusion polymerase in combination with the primers used here would result in detection of different community profiles. More importantly, it would show whether the Phusion polymerase results in the improved detection of high G+C species, whose detection was not enhanced with the methods tested and described here.

Detection frequencies of sequences assessed were not only grouped by library and therefore by primer, but also by patient. It was found that patients could be clustered according to subject-specific microbiota, specifically dominance of the microbiota by *Prevotella* or *Lactobacillus* species. This phenomenon had been noted previously (Chhour et al. 2005; Gross et al. 2010). This suggests that a useful future study would be the characterisation of a high number of patients to see if all patients can be grouped according to prevalence of certain species or genera and if different preventative and/or curative treatments for each specific group could consequently be deduced. In fact, the rationale for re-visiting the six patient samples using pyrosequencing with one primer set was to assess the comparability of next generation sequencing to established community profiling methods, to see whether pyrosequencing would be at least as effective in characterising the caries microbiota. If it was found to be equally as accurate, this method could help achieve the analysis of the microbiota in greater depth, thereby raising the likelihood of detecting rare species or species/patient-associated characteristics, such as subject-specific clusterings of genera. Assays could then be developed that might facilitate determination of a patient's 'caries group' by taking a sample prior to preliminary treatment that removes only the most degraded tissue. Once the concept of patient groups has been established, probing would only target the prevalent genera characterising each patient group, thereby reducing the amount of overall analysis needed. Ultimately, the lesion could be treated in a specific, personalised manner, resulting in overall minimally invasive treatments and reduction in chances of secondary caries. Such treatments might consist of use of antimicrobials specifically targeting the prevalent genera or families for that patient group, or identification and subsequent targeting of specific coaggregation bonds for patient group-specific bacteria. Furthermore, research assessing the patient-group specific metagenomes might identify disease-associated genes whose expression might be blocked by specific inhibitors. Indeed, a distinct example highlighting the need to not only gain knowledge about phylotypes, but also metagenomic and phenotypic characteristics, is the case of *Veillonella*. A metagenomic study demonstrated that caries-associated *Veillonella parvula* contained genes that the same species detected in caries-free patients did not (Belda-Ferre et al. 2012). This raises the question whether this also applies to different species, such as the predominant genera found in carious lesions, or the *Selenomonas* variants

detected in this study that were not abundant but showed distinct genetic polymorphisms on the 16S rDNA.

Overall, pyrosequencing resulted in the detection of a greater richness of taxa, a phenomenon generally observed when re-analysing habitats that have been sampled and studied using established methodologies (Santos et al. 2011; Griffen et al. 2012). Some species belonging to the phyla SR1 and *Chloroflexi* that make up the so-called rare biosphere were found, whose detection potential is increased using pyrosequencing due to the number of sequences returned from a single sequencing event (Keijser et al. 2008; Pedros-Alio 2012). Furthermore, higher Good's coverage rates in comparison to the Sanger-based analysis were noted. However, some genera were detected only with the PCR / cloning / sequencing approach. Moreover, at least ten species were found using culture, which 454 sequencing did not detect. At the same time CatchAll, which is arguably better suited for the analysis of samples that include many singletons, indicated that coverage rates were relatively low, with the majority being around or below 50 %, suggesting a far higher sampling effort is required to truly attain a data set representing the species that are present and their proportions (Diaz et al. 2012).

The debate as to whether *Actinobacteria* are underrepresented in molecular analyses, or, in fact, overrepresented in culture, was highlighted in both Chapters 3 and 4. Several experiments could help clear up any uncertainty regarding this issue. Firstly, Schwientek et al. (2011) reported problems with whole genome sequencing using the 454 system due to high G+C content and suggested use of an additive in the emulsion PCR reaction. Use of additives or adapted PCR protocols for the initial amplification, as well as the emulsion PCR, is certainly one area that should be empirically evaluated. Secondly, proportions of *Actinobacteria* were closest to those of a library using a different primer to what was used in the initial amplification for pyrosequencing reactions. This indicates that the methodology alone resulted in enhanced detection of *Actinobacteria*. Use of different forward primers, especially primer 61F, which resulted in the highest detection rate of the phylum *Actinobacteria* in Chapter 3, would be recommended if the same samples were to be re-analysed as a consequence of the finding that the forward primers seem to greatly influence detection rates on all phylogenetic levels. Furthermore, it might be beneficial to repeat experiments using

mock communities consisting of various low and high G+C species and the different forward primers using 454 sequencing, to assess the effect of the primers on accuracy of classification. This would give an indicator as to which and how many primers are needed to achieve a good representation of the present microbiota, including the high G+C species, potentially reducing time and cost spent on analysis in the long run. In addition, the use of FISH probes for the phylum *Actinobacteria* or its genera and subsequent visual evaluation of proportions found in either mock communities and/or actual patient samples would be valuable (Sim et al. 2012).

Indeed, FISH and the resulting enhanced understanding concerning the localisation of the various bacterial species in the caries lesion and the (improved) awareness of whether and/or how these species interact with each other and/or their environment is a further area that could potentially influence current preventative and/or treatment methods. It is known, for example, that colonisers of the developing plaque biofilm rely on coaggregation for the sequential adhesion. Some of these bacteria have several potential coaggregation partners utilising various protein adhesins and carbohydrate receptors that display different characteristics such as heat stability and protease sensitivity, while other species, such as *Capnocytophaga gingivalis*, have a limited coaggregation profile, requiring specific partners (Kolenbrander et al. 2006). If the visual investigation of the prevalent species in carious lesions found very specific key interactions of species driving lesion progression this would be an obvious target for the development of a coaggregation-inhibitor or a compound targeting any symbiotic networks, resulting in either preventative or treatment options. Consequently, methodologies potentially affecting and/or improving detection of these species and their (co-)localisation in the caries lesion were assessed in this study.

Probes to facilitate analysis of the prevalence and localisation of chosen genera and species in relation to all other microorganisms through the lesion were successfully developed and validated. No published study to date has attempted FISH on dentine from caries lesions using specific oligonucleotide probes targeting the prevalent taxa. FISH was performed successfully on excavated debris from bisected lesions hybridised with the universal probe EUB338, showing viable bacteria of various morphotypes, such as rods, cocci and filaments, confirming the potential to visualise bacteria in such samples, as has been reported previously (Banerjee et al. 2002). In fact, the method

used was developed for the quantification of bacteria in four distinct regions through a coronal carious lesion and the authors of the study suggested that development and application of specific probes to investigate the different layers of the caries lesion could result ultimately in improved clinical treatments (Banerjee et al. 2002). However, results of FISH on bisected teeth as well as excavated material, demonstrated one of the challenges of this technique, namely the difficulty of enabling the FISH probe to reach and penetrate the target cell and to allow specific hybridisation. Furthermore, the simultaneous staining of cells using the universal probe and a *Prevotella*-specific probe (multiplex FISH) resulted in successful visualisation of both target cells. A further problem was encountered, however, during multiplex FISH in that unspecific binding of probe to tooth debris was seen. Although the introduction of a pre-hybridisation step to block any unspecific binding sites, together with adjustments in the microscopy settings to avoid bleed-through, improved experimental outcomes, multiplex FISH was only successful in one instance. Both examples illustrate that further developmental work for the technique is needed. Solutions for the required modifications of the hybridisation protocol could simply consist of a change in incubation times and/or temperatures as well as the addition of one or several enzymes facilitating entry of probes into the fixed bacterial cells (Thurnheer et al. 2001; Quevedo et al. 2011). A more radical solution could be to try an alternative fixation protocol, such as, for example, paraformaldehyde fixation (Al-Ahmad et al. 2007; Hannig et al. 2007). A further method, CARD-FISH, the catalysed reporter deposition FISH (Pernthaler et al. 2002), is described as suitable for aquatic habitats; however, it may be worth applying this method on caries samples, since it is especially fitting for samples suffering from low signal intensity and high background fluorescence. A further potential strategy to improve signal intensity include the use of locked nucleic acids (LNAs). In this method LNA-incorporated oligonucleotide probes, in which two to four of these RNA derivatives are included in the oligonucleotide probe, have been found to enhance fluorescence intensity of probes by up to 22-fold (Kubota et al. 2006).

Though it could be demonstrated that the FISH protocol can be applied successfully for the direct visualisation of bacteria in caries lesions on bisected teeth, several technique and sample related issues remain. For example, due to not being able to see the staining of the tooth structure, indicating the different zones of lesion progression, it

was not possible to assess to which section the fluorescing cells belonged. If the aim was to investigate the bacteria in the advancing front only, it would be vital to ascertain how observed bacterial species correspond to staining of the tooth matrix. Furthermore, in the instance where cells could be visualised directly on the tooth, clumping of cells was observed, which raised the question as to whether hybridisation would be equally effective on the bacterial smears or excavated samples where cells were dispersed during handling. Moreover, it is not clear if sufficiently reproducible data can be extracted from such clumps, with regards to co-localisation and potential coaggregation, for the development of treatment protocols.

Further analysis methods such as metagenomic assessment of the regions (taken from the other half of the bisected tooth, to directly compare phylogenetic results with metagenomics), might give an indication whether species in the advancing front possess genes that species not associated with the disease progression do not. The difficulty with this approach is the inter-patient variability, though this may be reduced if patients can be classified in groups. A solution for this problem of identifying so-called biomarkers was presented by Segata et al. (2011) who presented the online interface Galaxy for the analysis of metagenomic data. Galaxy determines features that are most likely to explain differences between classes, such as genes or functions, by coupling standard tests for statistical significance with additional tests encoding biological consistency and effect relevance. This means the interface detects not only qualitative differences between classes (present/not present), but also quantitative features such as the estimation of the size of significant differences based on abundance patterns, creating a tool to rank biomarkers with regards to their biological aspect (Segata et al. 2011).

As described above in this chapter, enormous advances are currently being made in microbial ecology, as a result of the introduction of metagenomic as well as metabolomic and transcriptomic methodologies. These are all applicable to dental caries and will allow the determination of not only the composition of the caries-associated microbiota, but its functional potential and actual activity at different stages of carious lesion progression. This will enable the identification of new molecular targets for preventive and therapeutic interventions.

Bibliography

Aas, J. A., A. L. Griffen, S. R. Dardis, A. M. Lee, I. Olsen, F. E. Dewhirst, E. J. Leys and B. J. Paster (2008). "Bacteria of dental caries in primary and permanent teeth in children and young adults." J Clin Microbiol **46**(4): 1407-1417.

Aas, J. A., B. J. Paster, L. N. Stokes, I. Olsen and F. E. Dewhirst (2005). "Defining the normal bacterial flora of the oral cavity." J Clin Microbiol **43**(11): 5721--5732.

Absolom, D. R., F. V. Lamberti, Z. Policova, W. Zingg, C. J. van Oss and A. W. Neumann (1983). "Surface thermodynamics of bacterial adhesion." Appl. Environ. Microbiol. **46**(1): 90-97.

Aguirre, A., L. A. Testa-Weintraub, J. A. Banderas, G. G. Haraszthy, M. S. Reddy and M. J. Levine (1993). "Sialochemistry: a diagnostic tool?" Crit. Rev. Oral Biol. Med. **4**(3-4): 343-350.

Ahn, J., L. Yang, B. J. Paster, I. Ganly, L. Morris, Z. Pei and R. B. Hayes (2011). "Oral microbiome profiles: 16S rRNA pyrosequencing and microarray assay comparison." PLoS One **6**(7): e22788.

Akpan, A. and R. Morgan (2002). "Oral candidiasis." Postgrad. Med. J. **78**(922): 455-459.

Al-Ahmad, A., M. Follo, A. C. Selzer, E. Hellwig, M. Hannig and C. Hannig (2009). "Bacterial colonization of enamel in situ investigated using fluorescence in situ hybridization." J. Med. Microbiol. **58**(Pt 10): 1359-1366.

Al-Ahmad, A., A. Wunder, T. M. Auschill, M. Follo, G. Braun, E. Hellwig and N. B. Arweiler (2007). "The *in vivo* dynamics of *Streptococcus spp*., *Actinomyces naeslundii*, *Fusobacterium nucleatum* and *Veillonella* spp. in dental plaque biofilm as analysed by five-colour multiplex fluorescence *in situ* hybridization." J. Med. Microbiol. **56**(Pt 5): 681-687.

Allison, D. G. (2003). "The biofilm matrix." Biofouling **19**(2): 139-150.

Almstahl, A. and M. Wikstrom (1999). "Oral microflora in subjects with reduced salivary secretion." J. Dent. Res. **78**(8): 1410-1416.

Amann, R. I., W. Ludwig and K. H. Schleifer (1995). "Phylogenetic identification and in situ detection of individual microbial cells without cultivation." Microbiol Rev **59**(1): 143-169.

Amjad, Z. and G. H. Nancollas (1979). "Effect of fluoride on the growth of hydroxyapatite and human dental enamel." Caries Res. **13**(5): 250-258.

Andersen, R. N., N. Ganeshkumar and P. E. Kolenbrander (1993). "Cloning of the Streptococcus gordonii PK488 gene, encoding an adhesin which mediates coaggregation with Actinomyces naeslundii PK606." Infect. Immun. **61**(3): 981-987.

Armougom, F. and D. Raoult (2009). "Exploring Microbial Diversity Using 16S rRNA High-Throughput Methods." Journal of Computer Science & Systems Biology(2): 074-092.

Arnold, W. H., S. Konopka, M. S. Kriwalsky and P. Gaengler (2003). "Morphological analysis and chemical content of natural dentin carious lesion zones." Ann. Anat. **185**(5): 419-424.

Ashelford, K. E., N. A. Chuzhanova, J. C. Fry, A. J. Jones and A. J. Weightman (2005). "At least 1 in 20 16S rRNA sequence records currently held in public repositories is estimated to contain substantial anomalies." Appl. Environ. Microbiol. **71**(12): 7724-7736.

Ashelford, K. E., N. A. Chuzhanova, J. C. Fry, A. J. Jones and A. J. Weightman (2006). "New screening software shows that most recent large 16S rRNA gene clone libraries contain chimeras." Appl. Environ. Microbiol. **72**(9): 5734-5741.

Bagaitkar, J., L. R. Williams, D. E. Renaud, M. R. Bemakanakere, M. Martin, D. A. Scott and D. R. Demuth (2009). "Tobacco-induced alterations to Porphyromonas gingivalis-host interactions." Environmental microbiology **11**(5): 1242-1253.

Bagg, J., T. W. MacFarlane, I. R. Poxton, A. J. Smith and S. Bagg (1999). Essentials of Microbiology for Dental Students. Oxford, Oxford University Press.

Baker, G. C., J. J. Smith and D. A. Cowan (2003). "Review and re-analysis of domain-specific 16S primers." J. Microbiol. Methods **55**(3): 541-555.

Banerjee, A. (1999). Applications of Scanning Microscopy in the Assessment of Dentine Caries and Methods for Its Removal PhD, University of London.

Banerjee, A. and T. F. Watson (2011). Pickard's Manual of Operative Dentistry, OUP Oxford.

Banerjee, A., M. Yasseri and M. Munson (2002). "A method for the detection and quantification of bacteria in human carious dentine using fluorescent in situ hybridisation." J Dent **30**(7-8): 359-363.

Baron, A. C., S. A. Gansky, M. I. Ryder and J. D. Featherstone (1999). "Cysteine protease inhibitory activity and levels of salivary cystatins in whole saliva of periodontally diseased patients." J. Periodontal Res. **34**(8): 437-444.

Barriuso, J., J. R. Valverde and R. P. Mellado (2011). "Estimation of bacterial diversity using next generation sequencing of 16S rDNA: a comparison of different workflows." BMC Bioinformatics **12**: 473.

Baskaran, N., R. P. Kandpal, A. K. Bhargava, M. W. Glynn, A. Bale and S. M. Weissman (1996). "Uniform amplification of a mixture of deoxyribonucleic acids with varying GC content." Genome Res. **6**(7): 633-638.

Beatman, L. H. (1933). "Studies on *Trichomonas Buccalis*." J. Dent. Res. **13**(5): 339-347.

Becker, M. R., B. J. Paster, E. J. Leys, M. L. Moeschberger, S. G. Kenyon, J. L. Galvin, S. K. Boches, F. E. Dewhirst and A. L. Griffen (2002). "Molecular analysis of bacterial species associated with childhood caries." J Clin Microbiol **40**(3): 1001-1009.

Behrens, S., C. Rühland, J. Inácio, H. Huber, Á. Fonseca, I. Spencer-Martins, B. M. Fuchs and R. Amann (2003). "In Situ Accessibility of Small-Subunit rRNA of Members of the Domains Bacteria, Archaea, and Eucarya to Cy3-Labeled Oligonucleotide Probes." Appl. Environ. Microbiol. **69**(3): 1748-1758.

Beighton, D., M. Al-Haboubi, M. Mantzourani, S. C. Gilbert, D. Clark, L. Zoitopoulos and J. E. Gallagher (2010). "Oral Bifidobacteria: caries-associated bacteria in older adults." J. Dent. Res. **89**(9): 970-974.

Belda-Ferre, P., L. D. Alcaraz, R. Cabrera-Rubio, H. Romero, A. Simon-Soro, M. Pignatelli and A. Mira (2012). "The oral metagenome in health and disease." The ISME journal **6**(1): 46-56.

Biesbrock, A. R., M. S. Reddy and M. J. Levine (1991). "Interaction of a salivary mucin-secretory immunoglobulin A complex with mucosal pathogens." Infect. Immun. **59**(10): 3492-3497.

Bik, E. M., C. D. Long, G. C. Armitage, P. Loomer, J. Emerson, E. F. Mongodin, K. E. Nelson, S. R. Gill, C. M. Fraser-Liggett and D. A. Relman (2010). "Bacterial diversity in the oral cavity of 10 healthy individuals." The ISME journal **4**(8): 962-974.

Boles, B. R., M. Thoendel and P. K. Singh (2004). "Self-generated diversity produces "insurance effects" in biofilm communities." Proc. Natl. Acad. Sci. U. S. A. **101**(47): 16630-16635.

Bookstein, R., C. C. Lai, H. To and W. H. Lee (1990). "PCR-based detection of a polymorphic BamHI site in intron 1 of the human retinoblastoma (RB) gene." Nucleic Acids Res. **18**(6): 1666.

Borch, T. S., M. Lobner, K. Bendtzen, P. Holmstrup and C. H. Nielsen (2009). "Decreased interleukin-2 responses to Fusobacterium nucleatum and Porphyromonas gingivalis in generalized aggressive periodontitis." J. Periodontol. **80**(5): 800-807.

Borneman, J., P. W. Skroch, K. M. O'Sullivan, J. A. Palus, N. G. Rumjanek, J. L. Jansen, J. Nienhuis and E. W. Triplett (1996). "Molecular microbial diversity of an agricultural soil in Wisconsin." Appl. Environ. Microbiol. **62**(6): 1935-1943.

Bosch, J. A., M. Turkenburg, K. Nazmi, E. C. I. Veerman, E. J. C. de Geus and A. V. N. Amerongen (2003). "Stress as a determinant of saliva-mediated adherence and coadherence of oral and nonoral microorganisms." Psychosom. Med. **65**(4): 604-612.

Branda, S. S., S. Vik, L. Friedman and R. Kolter (2005). "Biofilms: the matrix revisited." Trends Microbiol. **13**(1): 20-26.

Bretz, W. A., P. M. Corby, T. C. Hart, S. Costa, M. Q. Coelho, R. J. Weyant, M. Robinson and N. J. Schork (2005). "Dental caries and microbial acid production in twins." Caries Res. **39**(3): 168-172.

Brown, L. R., S. Dreizen, S. Handler and D. A. Johnston (1975). "Effect of radiation-induced xerostomia on human oral microflora." J. Dent. Res. **54**(4): 740-750.

Bru, D., F. Martin-Laurent and L. Philippot (2008). "Quantification of the detrimental effect of a single primer-template mismatch by real-time PCR using the 16S rRNA gene as an example." Appl. Environ. Microbiol. **74**(5): 1660-1663.

Burke, F. M. and D. Y. Samarawickrama (1995). "Progressive changes in the pulpo-dentinal complex and their clinical consequences." Gerodontology **12**(12): 57-66.

Butler, W. T. (1995). "Dentin matrix proteins and dentinogenesis." Connective tissue research **33**(1-3): 59-65.

Byun, R., M. A. Nadkarni, K. L. Chhour, F. E. Martin, N. A. Jacques and N. Hunter (2004). "Quantitative analysis of diverse Lactobacillus species present in advanced dental caries." J Clin Microbiol **42**(7): 3128-3136.

Chang, W.-S., M. van de Mortel, L. Nielsen, G. N. de Guzman, X. Li and L. J. Halverson (2007). "Alginate production by Pseudomonas putida creates a hydrated microenvironment and contributes to biofilm architecture and stress tolerance under water-limiting conditions." J. Bacteriol.: 8290-8299.

Chapple, I. L. and J. Hamburger (2000). "The significance of oral health in HIV disease." Sex. Transm. Infect. **76**(4): 236-243.

Chhour, K. L., M. A. Nadkarni, R. Byun, F. E. Martin, N. A. Jacques and N. Hunter (2005). "Molecular analysis of microbial diversity in advanced caries." J Clin Microbiol **43**(2): 843-849.

Choi, B. K., B. J. Paster, F. E. Dewhirst and U. B. Gobel (1994). "Diversity of cultivable and uncultivable oral spirochetes from a patient with severe destructive periodontitis." Infect. Immun. **62**(5): 1889-1895.

Christensen, B. E., J. Kjosbakken and O. Smidsrod (1985). "Partial Chemical and Physical Characterization of Two Extracellular Polysaccharides Produced by Marine,

Periphytic Pseudomonas sp. Strain NCMB 2021." Appl. Environ. Microbiol. **50**(4): 837-845.

Clarke, J. K. (1924). "On the bacterial factor in the aetiology of dental caries." Br J Exp Pathol **5**: 141-147.

Clarkson, B. H., S. R. Chang and G. R. Holland (1998). "Phosphoprotein analysis of sequential extracts of human dentin and the determination of the subsequent remineralization potential of these dentin matrices." Caries Res. **32**(5): 357-364.

Colwell, R. K. (2009). "EstimateS: Statistical estimation of species richness and shared species from samples. Version 8.2.", from User's Guide and application published at: http://purl.oclc.org/estimates.

Conover, M. S., M. Mishra and R. Deora (2011). "Extracellular DNA Is Essential for Maintaining *Bordetella* Biofilm Integrity on Abiotic Surfaces and in the Upper Respiratory Tract of Mice." PLoS One **6**(2): e16861.

Correia, F. F., J. M. DiRienzo, T. L. McKay and B. Rosan (1996). "scbA from *Streptococcus crista* CC5A: an atypical member of the lraI gene family." Infect. Immun. **64**(6): 2114-2121.

Costello, E. K., C. L. Lauber, M. Hamady, N. Fierer, J. I. Gordon and R. Knight (2009). "Bacterial community variation in human body habitats across space and time." Science **326**(5960): 1694-1697.

Costerton, J. W., Z. Lewandowski, D. E. Caldwell, D. R. Korber and H. M. Lappin-Scott (1995). "Microbial biofilms." Annu. Rev. Microbiol. **49**: 711-745.

Costerton, J. W., Z. Lewandowski, D. DeBeer, D. Caldwell, D. Korber and G. James (1994). "Biofilms, the customized microniche." J. Bacteriol. **176**(8): 2137-2142.

Das, T., P. K. Sharma, H. J. Busscher, H. C. van der Mei and B. P. Krom (2010). "Role of Extracellular DNA in Initial Bacterial Adhesion and Surface Aggregation." Appl. Environ. Microbiol. **76**(10): 3405-3408.

Davey, M. E. and G. A. O'Toole (2000). "Microbial biofilms: from ecology to molecular genetics." Microbiol. Mol. Biol. Rev. **64**(4): 847-867.

Davies, D. G., M. R. Parsek, J. P. Pearson, B. H. Iglewski, J. W. Costerton and E. P. Greenberg (1998). "The involvement of cell-to-cell signals in the development of a bacterial biofilm." Science **280**: 295-297.

Dayan, S., P. Stashenko, R. Niederman and T. S. Kupper (2004). "Oral epithelial overexpression of IL-1alpha causes periodontal disease." J. Dent. Res. **83**(10): 786-790.

de Boer, R., R. Peters, S. Gierveld, T. Schuurman, M. Kooistra-Smid and P. Savelkoul (2010). "Improved detection of microbial DNA after bead-beating before DNA isolation." J. Microbiol. Methods **80**(2): 209-211.

de Lillo, A., F. P. Ashley, R. M. Palmer, M. A. Munson, L. Kyriacou, A. J. Weightman and W. G. Wade (2006). "Novel subgingival bacterial phylotypes detected using multiple universal polymerase chain reaction primer sets." Oral Microbiol. Immunol. **21**(1): 61-68.

de Villiers, E. M., H. Weidauer, H. Otto and H. zur Hausen (1985). "Papillomavirus DNA in human tongue carcinomas." Int. J. Cancer **36**(5): 575-578.

Deeley, K., A. Letra, E. K. Rose, C. A. Brandon, J. M. Resick, M. L. Marazita and A. R. Vieira (2008). "Possible association of amelogenin to high caries experience in a Guatemalan-Mayan population." Caries Res. **42**(1): 8-13.

Devereux, R. and G. W. Mundfrom (1994). "A phylogenetic tree of 16S rRNA sequences from sulfate-reducing bacteria in a sandy marine sediment." Appl. Environ. Microbiol. **60**(9): 3437-3439.

Dewhirst, F. E., T. Chen, J. Izard, B. J. Paster, A. C. Tanner, W. H. Yu, A. Lakshmanan and W. G. Wade (2010). "The human oral microbiome." J. Bacteriol. **192**(19): 5002-5017.

Di Luca, D., P. Mirandola, T. Ravaioli, R. Dolcetti, A. Frigatti, P. Bovenzi, L. Sighinolfi, P. Monini and E. Cassai (1995). "Human herpesviruses 6 and 7 in salivary glands and shedding in saliva of healthy and human immunodeficiency virus positive individuals." J. Med. Virol. **45**(4): 462-468.

Diaz, P. I., N. I. Chalmers, A. H. Rickard, C. Kong, C. L. Milburn, R. J. Palmer and P. E. Kolenbrander (2006). "Molecular characterization of subject-specific oral microflora during initial colonization of enamel." Appl. Environ. Microbiol. **72**(4): 2837-2848.

Diaz, P. I., A. K. Dupuy, L. Abusleme, B. Reese, C. Obergfell, L. Choquette, A. Dongari-Bagtzoglou, D. E. Peterson, E. Terzi and L. D. Strausbaugh (2012). "Using high throughput sequencing to explore the biodiversity in oral bacterial communities." Molecular oral microbiology **27**(3): 182-201.

Dige, I., H. Nilsson, M. Kilian and B. Nyvad (2007). "*In situ* identification of *streptococci* and other bacteria in initial dental biofilm by confocal laser scanning microscopy and fluorescence *in situ* hybridization." Eur. J. Oral Sci. **115**(6): 459-467.

Dirks, O. B. (1966). "Posteruptive Changes in Dental Enamel." J. Dent. Res. **45**(3): 503-511.

Do, T., K. A. Jolley, M. C. Maiden, S. C. Gilbert, D. Clark, W. G. Wade and D. Beighton (2009). "Population structure of Streptococcus oralis." Microbiology **155**(Pt 8): 2593-2602.

Don, R. H., P. T. Cox, B. J. Wainwright, K. Baker and J. S. Mattick (1991). "'Touchdown' PCR to circumvent spurious priming during gene amplification." Nucleic Acids Res. **19**(14): 4008.

Dowker, S. E., J. C. Elliott, G. R. Davis and H. S. Wassif (2003). "Longitudinal study of the three-dimensional development of subsurface enamel lesions during in vitro demineralisation." Caries Res. **37**(4): 237-245.

Duenas-Chasco, M. T., M. A. Rodriguez-Carvajal, P. Tejero-Mateo, J. L. Espartero, A. Irastorza-Iribas and A. M. Gil-Serrano (1998). "Structural analysis of the exopolysaccharides produced by Lactobacillus spp. G-77." Carbohydr. Res. **307**(1-2): 125-133.

Dumas, C., A. Champagne and M. C. Lavoie (1987). "Proteolytic activity of bacteria isolated from the oral cavities of BALB/c mice toward salivary proteins." J. Dent. Res. **66**(1): 62-64.

Dutton, C. M., C. Paynton and S. S. Sommer (1993). "General method for amplifying regions of very high G+C content." Nucleic Acids Res. **21**(12): 2953-2954.

Dymock, D., A. J. Weightman, C. Scully and W. G. Wade (1996). "Molecular analysis of microflora associated with dentoalveolar abscesses." J Clin Microbiol **34**(3): 537-542.

Eckburg, P. B., E. M. Bik, C. N. Bernstein, E. Purdom, L. Dethlefsen, M. Sargent, S. R. Gill, K. E. Nelson and D. A. Relman (2005). "Diversity of the human intestinal microbial flora." Science **308**(5728): 1635-1638.

Edwards, U., T. Rogall, H. Blocker, M. Emde and E. C. Bottger (1989). "Isolation and direct complete nucleotide determination of entire genes. Characterization of a gene coding for 16S ribosomal RNA." Nucleic Acids Res. **17**(19): 7843-7853.

Egland, P. G., L. D. D and P. E. Kolenbrander (2001). "Identification of independent Streptococcus gordonii SspA and SspB functions in coaggregation with Actinomyces naeslundii." Infect. Immun. **69**(12): 7512-7516.

Farris, M. H. and J. B. Olson (2007). "Detection of Actinobacteria cultivated from environmental samples reveals bias in universal primers." Lett. Appl. Microbiol. **45**(4): 376-381.

Featherstone, J. D. (2008). "Dental caries: a dynamic disease process." Australian dental journal **53**(3): 286-291.

Fejerskov, O. and E. A. Kidd, Eds. (2003). Dental Caries, Blackwell Munksgaard.

Fincham, A. G., J. Moradian-Oldak and J. P. Simmer (1999). "The structural biology of the developing dental enamel matrix." J. Struct. Biol. **126**(3): 270-299.

Fitzgerald, R. J., H. V. Jordan and H. O. Archard (1966). "Dental caries in gnotobiotic rats infected with a variety of Lactobacillus acidophilus." Arch. Oral Biol. **11**(5): 473-476.

Foster, J. S. and P. E. Kolenbrander (2004). "Development of a multispecies oral bacterial community in a saliva-conditioned flow cell." Appl. Environ. Microbiol. **70**(7): 4340-4348.

Fox, G. E., J. D. Wisotzkey and P. Jurtshuk, Jr. (1992). "How close is close: 16S rRNA sequence identity may not be sufficient to guarantee species identity." International journal of systematic bacteriology **42**(1): 166-170.

Frandsen, E. V., V. Pedrazzoli and M. Kilian (1991). "Ecology of viridans streptococci in the oral cavity and pharynx." Oral Microbiol. Immunol. **6**(3): 129-133.

Frank, J. A., C. I. Reich, S. Sharma, J. S. Weisbaum, B. A. Wilson and G. J. Olsen (2008). "Critical evaluation of two primers commonly used for amplification of bacterial 16S rRNA genes." Appl. Environ. Microbiol. **74**(8): 2461-2470.

Frey, U. H., H. S. Bachmann, J. Peters and W. Siffert (2008). "PCR-amplification of GC-rich regions: 'slowdown PCR'." Nature protocols **3**(8): 1312-1317.

Fuchs, B. M., G. Wallner, W. Beisker, I. Schwippl, W. Ludwig and R. Amann (1998). "Flow cytometric analysis of the in situ accessibility of Escherichia coli 16S rRNA for fluorescently labeled oligonucleotide probes." Appl. Environ. Microbiol. **64**(12): 4973-4982.

Ganeshkumar, N., M. Song and B. C. McBride (1988). "Cloning of a Streptococcus sanguis adhesin which mediates binding to saliva-coated hydroxyapatite." Infect. Immun. **56**(5): 1150-1157.

Garberoglio, R. and M. Brännström (1976). "Scanning electron microscopic investigation of human dentinal tubules." Arch. Oral Biol. **21**(6): 355-362.

Gart, J. J. (1982). "Rarefaction and Taxonomic Diversity." Biometrics **38**: 235-241.

Ghabanchi, J., M. Zibaei, M. D. Afkar and A. H. Sarbazie (2010). "Prevalence of oral Entamoeba gingivalis and Trichomonas tenax in patients with periodontal disease and healthy population in Shiraz, southern Iran." Indian journal of dental research : official publication of Indian Society for Dental Research **21**(1): 89-91.

Gibbons, R. J. (1989). "Bacterial adhesion to oral tissues: a model for infectous diseases." J. Dent. Res. **68**: 750-760.

Gibbons, R. J. and D. I. Hay (1988). "Human salivary acidic proline-rich proteins and statherin promote the attachment of Actinomyces viscosus LY7 to apatitic surfaces." Infect. Immun. **56**(2): 439-445.

Gibbons, R. J., D. I. Hay, J. O. Cisar and W. B. Clark (1988). "Adsorbed salivary proline-rich protein 1 and statherin: receptors for type 1 fimbriae of Actinomyces viscosus T14V-J1 on apatitic surfaces." Infect. Immun. **56**(11): 2990-2993.

Gibbons, R. J., D. I. Hay and D. H. Schlesinger (1991). "Delineation of a segment of adsorbed salivary acidic proline-rich proteins which promotes adhesion of *Streptococcus gordonii* to apatitic surfaces." Infect. Immun. **59**(9): 2948-2954.

Gilbert, P., J. R. Das, M. V. Jones and D. G. Allison (2001). "Assessment of resistance towards biocides following the attachment of micro-organisms to, and growth on, surfaces." J. Appl. Microbiol. **91**(2): 248-254.

Giovannoni, S. J., T. B. Britschgi, C. L. Moyer and K. G. Field (1990). "Genetic diversity in Sargasso Sea bacterioplankton." Nature **345**(6270): 60-63.

Good, I. J. (1953). "The Population Frequencies of Species and the Estimation of Population Parameters." Biometrika **40**(3-4): 237-264.

Gothefors, L. and S. Marklund (1975). "Lactoperoxidase activity in human milk and in saliva of newborn infants." Infect. Immun. **11**(6): 1210-1215.

Gray, J. P. and R. P. Herwig (1996). "Phylogenetic analysis of the bacterial communities in marine sediments." Appl. Environ. Microbiol. **62**(11): 4049-4059.

Grice, E. A. and J. A. Segre (2011). "The skin microbiome." Nat. Rev. Microbiol. **9**(4): 244-253.

Griffen, A. L., C. J. Beall, J. H. Campbell, N. D. Firestone, P. S. Kumar, Z. K. Yang, M. Podar and E. J. Leys (2012). "Distinct and complex bacterial profiles in human periodontitis and health revealed by 16S pyrosequencing." The ISME journal **6**(6): 1176-1185.

Gross, E. L., E. J. Leys, S. R. Gasparovich, N. D. Firestone, J. A. Schwartzbaum, D. A. Janies, K. Asnani and A. L. Griffen (2010). "Bacterial 16S sequence analysis of severe caries in young permanent teeth." J Clin Microbiol **48**(11): 4121-4128.

Gruter, M., B. R. Leeflang, J. Kuiper, J. P. Kamerling and J. F. Vliegenthart (1993). "Structural characterisation of the exopolysaccharide produced by Lactobacillus delbruckii subspecies bulgaricus rr grown in skimmed milk." Carbohydr. Res. **239**: 209-226.

Guggenheim, M., S. Shapiro, R. Gm?r and B. Guggenheim (2001). "Spatial arrangements and associative behavior of species in an in vitro oral biofilm model." Appl. Environ. Microbiol. **67**(3): 1343-1350.

Guobis, Z., V. Kareiviene, N. Baseviciene, P. Paipaliene, I. Niedzelskiene, G. Sabalys, R. Kubilius and A. Gervickas (2011). "Microflora of the oral cavity in patients with xerostomia." Medicina (Kaunas) **47**(12): 646-651.

Haffajee, A. D. and S. S. Socransky (2005). "Microbiology of periodontal diseases: introduction." Periodontol 2000 **38**: 9-12.

Haffajee, A. D., S. S. Socransky, M. R. Patel and X. Song (2008). "Microbial complexes in supragingival plaque." Oral Microbiol. Immunol. **23**(3): 196-205.

Hall, T. (2011). "Bioedit." Version 7.1.3. from http://www.mbio.ncsu.edu/bioedit/bioedit.html.

Hamada, S. and H. D. Slade (1980). "Biology, immunology, and cariogenicity of Streptococcus mutans." Microbiol Rev **44**(2): 331-384.

Hanage, W. P., C. Fraser and B. G. Spratt (2005). "Fuzzy species among recombinogenic bacteria." BMC Biol **3**: 6.

Hanage, W. P., C. Fraser and B. G. Spratt (2006). "Sequences, sequence clusters and bacterial species." Philos. Trans. R. Soc. Lond. B Biol. Sci. **361**(1475): 1917-1927.

Hannig, C. and M. Hannig (2009). "The oral cavity--a key system to understand substratum-dependent bioadhesion on solid surfaces in man." Clin Oral Investig **13**(2): 123-139.

Hannig, C., M. Hannig and T. Attin (2005). "Enzymes in the acquired enamel pellicle." Eur. J. Oral Sci. **113**(1): 2-13.

Hannig, C., M. Hannig, O. Rehmer, G. Braun, E. Hellwig and A. Al-Ahmad (2007). "Fluorescence microscopic visualization and quantification of initial bacterial colonization on enamel *in situ*." Arch. Oral Biol.: 1048-1056.

Hannig, M. and A. Joiner, Eds. (2006). The structure, function and properties of the acquired pellicle. The teeth and their environment, Monographs in Oral Science.

Hansen, M. C., T. Tolker-Nielsen, M. Givskov and S. Molin (1998). "Biased 16S rDNA PCR amplification caused by interference from DNA flanking the template region." FEMS Microbiol. Ecol. **26**(2): 141-149.

Haukioja, A., E. Soderling and J. Tenovuo (2008). "Acid production from sugars and sugar alcohols by probiotic lactobacilli and bifidobacteria in vitro." Caries Res. **42**(6): 449-453.

Hendolin, P. H., L. Paulin and J. Ylikoski (2000). "Clinically applicable multiplex PCR for four middle ear pathogens." J Clin Microbiol **38**(1): 125-132.

Henke, W., K. Herdel, K. Jung, D. Schnorr and S. A. Loening (1997). "Betaine improves the PCR amplification of GC-rich DNA sequences." Nucleic Acids Res. **25**(19): 3957-3958.

Henssge, U., T. Do, S. C. Gilbert, S. Cox, D. Clark, C. Wickstrom, A. J. Ligtenberg, D. R. Radford and D. Beighton (2011). "Application of MLST and pilus gene sequence comparisons to investigate the population structures of Actinomyces naeslundii and Actinomyces oris." PLoS One **6**(6): e21430.

Henssge, U., T. Do, D. R. Radford, S. C. Gilbert, D. Clark and D. Beighton (2009). "Emended description of Actinomyces naeslundii and descriptions of Actinomyces oris sp. nov. and Actinomyces johnsonii sp. nov., previously identified as Actinomyces naeslundii genospecies 1, 2 and WVA 963." Int. J. Syst. Evol. Microbiol. **59**(Pt 3): 509-516.

Hill, J. E., J. R. Town and S. M. Hemmingsen (2006). "Improved template representation in cpn60 polymerase chain reaction (PCR) product libraries generated from complex templates by application of a specific mixture of PCR primers." Environmental microbiology **8**(4): 741-746.

Horakova, H., I. Polakovicova, G. M. Shaik, J. Eitler, V. Bugajev, L. Draberova and P. Draber (2011). "1,2-propanediol-trehalose mixture as a potent quantitative real-time PCR enhancer." BMC Biotechnol. **11**: 41.

Horiuchi, M., J. Washio, H. Mayanagi and N. Takahashi (2009). "Transient acid-impairment of growth ability of oral Streptococcus, Actinomyces, and Lactobacillus: a possible ecological determinant in dental plaque." Oral Microbiol. Immunol. **24**(4): 319-324.

Horz, H. P. and G. Conrads (2011). "Methanogenic Archaea and oral infections - ways to unravel the black box." Journal of oral microbiology **3**.

Howe, P. R. (1917). "A Study of the Microorganisms of Dental Caries." The Journal of medical research **36**(3): 481-492 485.

Huew, R., P. Waterhouse, P. Moynihan, S. Kometa and A. Maguire (2012). "Dental caries and its association with diet and dental erosion in Libyan schoolchildren."

International journal of paediatric dentistry / the British Paedodontic Society [and] the International Association of Dentistry for Children **22**(1): 68-76.

Hughes, S. and A. Moody, Eds. (2007). PCR. The methods express series, Scion Publishing Ltd.

Humphrey, S. P. and R. T. Williamson (2001). "A review of saliva: Normal composition, flow, and function." The Journal of prosthetic dentistry **85**(2): 162-169.

Hung, T., K. Mak and K. Fong (1990). "A specificity enhancer for polymerase chain reaction." Nucleic Acids Res. **18**(16): 4953.

Janda, J. M. and S. L. Abbott (2007). "16S rRNA gene sequencing for bacterial identification in the diagnostic laboratory: pluses, perils, and pitfalls." J Clin Microbiol **45**(9): 2761-2764.

Jaramillo, A., R. M. Arce, D. Herrera, M. Betancourth, J. E. Botero and A. Contreras (2005). "Clinical and microbiological characterization of periodontal abscesses." J Clin Periodontol **32**(12): 1213-1218.

Jay, P. (1938). "Lactobacillus Acidophilus and Dental Caries." American journal of public health and the nation's health **28**(6): 759-761.

Jenkinson, H. F. and R. J. Lamont (2005). "Oral microbial communities in sickness and in health." Trends Microbiol. **13**(12): 589-595.

Johnson, S. A., P. A. Goddard, C. Iliffe, B. Timmins, A. H. Rickard, G. Robson and P. S. Handley (2002). "Comparative susceptibility of resident and transient hand bacteria to para-chloro-meta-xylenol and triclosan." J. Appl. Microbiol. **93**(2): 336-344.

Jong, M. H. D. and J. S. V. der Hoeven (1987). "The growth of oral bacteria on saliva." J. Dent. Res. **66**(2): 498-505.

Jung, D. J., A. Al-Ahmad, M. Follo, B. Spitzmuller, W. Hoth-Hannig, M. Hannig and C. Hannig (2010). "Visualization of initial bacterial colonization on dentine and enamel in situ." J. Microbiol. Methods **81**(2): 166-174.

Kanasi, E., F. E. Dewhirst, N. I. Chalmers, R. Kent, Jr., A. Moore, C. V. Hughes, N. Pradhan, C. Y. Loo and A. C. Tanner (2010). "Clonal analysis of the microbiota of severe early childhood caries." Caries Res. **44**(5): 485-497.

Kaufman, J. and J. M. DiRienzo (1989). "Isolation of a corncob (coaggregation) receptor polypeptide from Fusobacterium nucleatum." Infect. Immun. **57**(2): 331-337.

Keijser, B. J., E. Zaura, S. M. Huse, J. M. van der Vossen, F. H. Schuren, R. C. Montijn, J. M. ten Cate and W. Crielaard (2008). "Pyrosequencing analysis of the oral microflora of healthy adults." J. Dent. Res. **87**(11): 1016-1020.

Khader, Y. S., A. S. Dauod, S. S. El-Qaderi, A. Alkafajei and W. Q. Batayha (2006). "Periodontal status of diabetics compared with nondiabetics: a meta-analysis." J Diabetes Complications **20**(1): 59-68.

Kidd, E. A. and O. Fejerskov (2004). "What constitutes dental caries? Histopathology of carious enamel and dentin related to the action of cariogenic biofilms." J. Dent. Res. **83 Spec No C**: C35-38.

Kidd, E. A., S. Joyston-Bechal and D. Beighton (1993). "Microbiological evaluation of caries activity during cavity preparation." Caries Res. **27**: 402-408.

Kinder, S. A. and S. C. Holt (1993). "Localization of the Fusobacterium nucleatum T18 adhesin activity mediating coaggregation with Porphyromonas gingivalis T22." J. Bacteriol. **175**(3): 840-850.

Kleinberg, I. (1961). "Studies on Dental Plaque. I. The Effect of Different Concentrations of Glucose on the pH of Dental Plaque in Vivo." J. Dent. Res. **40**(6): 1087-1111.

Kleinberg, I. (2002). "A mixed-bacteria ecological approach to understanding the role of the oral bacteria in dental caries causation: an alternative to Streptococcus mutans and the specific-plaque hypothesis." Crit. Rev. Oral Biol. Med. **13**(2): 108-125.

Kohler, B., B. M. Pettersson and D. Bratthall (1981). "*Streptococcus mutans* in plaque and saliva and the development of caries." Scand J Dent Res **89**(1): 19-25.

Kolenbrander, P. E. (2000). "Oral microbial communities: biofilms, interactions, and genetic systems." Annual Reviews in Microbiology **54**: 413-437.

Kolenbrander, P. E. and R. N. Andersen (1986). "Multigeneric aggregations among oral bacteria: a network of independent cell-to-cell interactions." J. Bacteriol. **168**(2): 851-859.

Kolenbrander, P. E. and R. N. Andersen (1988). "Intergeneric rosettes: sequestered surface recognition among human periodontal bacteria." Appl. Environ. Microbiol. **54**(4): 1046-1050.

Kolenbrander, P. E., R. N. Andersen, D. S. Blehert, P. G. Egland, J. S. Foster and R. J. Palmer (2002). "Communication among oral bacteria." Microbiol. Mol. Biol. Rev. **66**(3): 486-505.

Kolenbrander, P. E., R. N. Andersen and L. V. Moore (1989). "Coaggregation of Fusobacterium nucleatum, Selenomonas flueggei, Selenomonas infelix, Selenomonas noxia, and Selenomonas sputigena with strains from 11 genera of oral bacteria." Infect. Immun. **57**(10): 3194-3203.

Kolenbrander, P. E., R. N. Andersen and L. V. Moore (1990). "Intrageneric coaggregation among strains of human oral bacteria: potential role in primary colonization of the tooth surface." Appl. Environ. Microbiol. **56**(12): 3890-3894.

Kolenbrander, P. E., P. G. Egland, P. I. Diaz and R. J. Palmer (2005). "Genome-genome interactions: bacterial communities in initial dental plaque." Trends Microbiol. **13**(1): 11-15.

Kolenbrander, P. E., N. Ganeshkumar, F. J. Cassels and C. V. Hughes (1993). "Coaggregation: specific adherence among human oral plaque bacteria." FASEB J. **7**(5): 406-413.

Kolenbrander, P. E., R. J. Palmer, Jr., A. H. Rickard, N. S. Jakubovics, N. I. Chalmers and P. I. Diaz (2006). "Bacterial interactions and successions during plaque development." Periodontol 2000 **42**: 47-79.

Krasse, B. (1966). "Human streptococci and experimental caries in hamsters." Arch. Oral Biol. **11**(4): 429-436.

Krekeler, C., H. Ziehr and J. Klein (1989). "Physical methods for characterization of microbial surfaces." Experientia **45**(11-12): 1047-1055.

Kroes, I., P. W. Lepp and D. A. Relman (1999). "Bacterial diversity within the human subgingival crevice." Proc. Natl. Acad. Sci. U. S. A. **96**(25): 14547-14552.

Kubota, K., A. Ohashi, H. Imachi and H. Harada (2006). "Improved in situ hybridization efficiency with locked-nucleic-acid-incorporated DNA probes." Appl. Environ. Microbiol. **72**(8): 5311-5317.

Lane, D. J. (1991). Nucleic acid techniques in bacterial systematics. Chichester, United Kingdom, John Wiley & Sons Ltd.

Lawrence, J. R., T. R. Neu and G. D. W. Swerhome (1998). "Application of multiple parameter imaging for the quantification of algal, bacterial and exopolymer components of microbial biofilms." J. Microbiol. Methods **32**: 253-261.

Lazarevic, V., K. Whiteson, S. Huse, D. Hernandez, L. Farinelli, M. Osteras, J. Schrenzel and P. Francois (2009). "Metagenomic study of the oral microbiota by Illumina high-throughput sequencing." J. Microbiol. Methods **79**(3): 266-271.

Lehner, T., J. Caldwell and R. Smith (1985). "Local passive immunization by monoclonal antibodies against streptococcal antigen I/II in the prevention of dental caries." Infect. Immun. **50**(3): 796-799.

Lehner, T., S. J. Challacombe and J. Caldwell (1975). "Immunological and bacteriological basis for vaccination against dental caries in rhesus monkeys." Nature **254**(5500): 517-520.

Lendenmann, U., J. Grogan and F. G. Oppenheim (2000). "Saliva and dental pellicle--a review." Adv Dent Res **14**: 22-28.

Leser, T. D., J. Z. Amenuvor, T. K. Jensen, R. H. Lindecrona, M. Boye and K. Moller (2002). "Culture-independent analysis of gut bacteria: the pig gastrointestinal tract microbiota revisited." Appl. Environ. Microbiol. **68**(2): 673-690.

Levine, M. J., M. S. Reddy, L. A. Tabak, R. E. Loomis, E. J. Bergey, P. C. Jones, R. E. Cohen, M. W. Stinson and I. Al-Hashimi (1987). "Structural aspects of salivary glycoproteins." J. Dent. Res. **66**(2): 436-441.

Li, J., E. J. Helmerhorst, C. W. Leone, R. F. Troxler, T. Yaskell, A. D. Haffajee, S. S. Socransky and F. G. Oppenheim (2004). "Identification of early microbial colonizers in human dental biofilm." J. Appl. Microbiol. **97**(6): 1311-1318.

Li, K., M. Bihan, S. Yooseph and B. A. Methe (2012). "Analyses of the Microbial Diversity across the Human Microbiome." PLoS One **7**(6): e32118.

Li, L., W. W. Hsiao, R. Nandakumar, S. M. Barbuto, E. F. Mongodin, B. J. Paster, C. M. Fraser-Liggett and A. F. Fouad (2010). "Analyzing endodontic infections by deep coverage pyrosequencing." J. Dent. Res. **89**(9): 980-984.

Ligtenberg, A. J., E. Walgreen-Weterings, E. C. Veerman, J. J. de Soet, J. de Graaff and A. V. Amerongen (1992). "Influence of saliva on aggregation and adherence of Streptococcus gordonii HG 222." Infect. Immun. **60**(9): 3878-3884.

Lima, K. C., L. T. Coelho, I. V. Pinheiro, I. N. Rocas and J. F. Siqueira, Jr. (2011). "Microbiota of dentinal caries as assessed by reverse-capture checkerboard analysis." Caries Res. **45**(1): 21-30.

Lin, S. Y., F. T. Shen and C. C. Young (2011). "Rapid detection and identification of the free-living nitrogen fixing genus Azospirillum by 16S rRNA-gene-targeted genus-specific primers." Antonie Leeuwenhoek **99**(4): 837-844.

Listgarten, M. A. (1986). "Pathogenesis of periodontitis." J Clin Periodontol **13**(5): 418-430.

Listgarten, M. A., H. E. Mayo and R. Tremblay (1975). "Development of dental plaque on epoxy resin crowns in man. A light and electron microscopic study." J. Periodontol. **46**(1): 10-26.

Llory, H., M. Gioanni, A. Dammron and R. M. Frank (1972). "Some Population Changes in Oral Anaerobic Microorganisms, Streptococcus-Mutans and Yeasts Following Irradiation of Salivary-Glands." Caries Res. **6**(4): 298-&.

Loe, H. (1993). "Periodontal disease. The sixth complication of diabetes mellitus." Diabetes care **16**(1): 329-334.

Loe, H., E. Theilade and S. B. Jensen (1965). "Experimental Gingivitis in Man." J. Periodontol. **36**: 177-187.

Loesche, W. J. (1976). "Chemotherapy of dental plaque infections." Oral sciences reviews **9**: 65-107.

Loesche, W. J. (1986). "Role of Streptococcus mutans in human dental decay." Microbiol Rev **50**(4): 353-380.

Loesche, W. J., J. Rowan, L. H. Straffon and P. J. Loos (1975). "Association of Streptococcus mutants with human dental decay." Infect. Immun. **11**(6): 1252-1260.

Loomis, R. E., A. Prakobphol, M. J. Levine, M. S. Reddy and P. C. Jones (1987). "Biochemical and biophysical comparison of two mucins from human submandibular-sublingual saliva." Arch. Biochem. Biophys. **258**(2): 452-464.

Lopez-Garcia, P. and D. Moreira (1999). "Metabolic symbiosis at the origin of eukaryotes." Trends Biochem. Sci. **24**(3): 88-93.

Lopez, N. J., L. Jara and C. Y. Valenzuela (2005). "Association of interleukin-1 polymorphisms with periodontal disease." J. Periodontol. **76**(2): 234-243.

Lourenco, S. V., T. P. Hussein, S. B. Bologna, A. M. Sipahi and M. M. Nico (2010). "Oral manifestations of inflammatory bowel disease: a review based on the observation of six cases." Journal of the European Academy of Dermatology and Venereology : JEADV **24**(2): 204-207.

Love, R. M. and H. F. Jenkinson (2002). "Invasion of dentinal tubules by oral bacteria." Critical Reviews in Oral Biology & Medicine **13**(2): 171-183.

Lynch, R. J., R. Navada and R. Walia (2004). "Low-levels of fluoride in plaque and saliva and their effects on the demineralisation and remineralisation of enamel; role of fluoride toothpastes." International dental journal **54**(5 Suppl 1): 304-309.

MacKay, B. J., L. Denepitiya, V. J. Iacono, S. B. Krost and J. J. Pollock (1984). "Growth-inhibitory and bactericidal effects of human parotid salivary histidine-rich polypeptides on Streptococcus mutans." Infect. Immun. **44**(3): 695-701.

Mager, D. L., L. A. Ximenez-Fyvie, A. D. Haffajee and S. S. Socransky (2003). "Distribution of selected bacterial species on intraoral surfaces." J Clin Periodontol **30**(7): 644-654.

Maidak, B. L., J. R. Cole, T. G. Lilburn, C. T. Parker, Jr., P. R. Saxman, R. J. Farris, G. M. Garrity, G. J. Olsen, T. M. Schmidt and J. M. Tiedje (2001). "The RDP-II (Ribosomal Database Project)." Nucleic Acids Res. **29**(1): 173-174.

Mann, A. B. and M. E. Dickinson (2006). "Nanomechanics, chemistry and structure at the enamel surface." Monogr. Oral Sci. **19**: 105-131.

Mantzourani, M., S. C. Gilbert, H. N. Sulong, E. C. Sheehy, S. Tank, M. Fenlon and D. Beighton (2009). "The isolation of bifidobacteria from occlusal carious lesions in children and adults." Caries Res. **43**(4): 308-313.

Marchesi, J. R., T. Sato, A. J. Weightman, T. A. Martin, J. C. Fry, S. J. Hiom, D. Dymock and W. G. Wade (1998). "Design and evaluation of useful bacterium-specific PCR primers that amplify genes coding for bacterial 16S rRNA." Appl. Environ. Microbiol. **64**(2): 795-799.

Marcotte, H. and M. C. Lavoie (1998). "Oral microbial ecology and the role of salivary immunoglobulin A." Microbiol. Mol. Biol. Rev. **62**(1): 71-109.

Mardis, E. R. (2008). "Next-generation DNA sequencing methods." Annu. Rev. Genomics Hum. Genet. **9**: 387-402.

Marmur, J. (1961). "A procedure for the isolation of deoxyribonucleic acid from micro-organisms." J. Mol. Biol. **3**(2): 208-IN201.

Marsh, P. D. (1994). "Microbial ecology of dental plaque and its significance in health and disease." Advances in Dental Research **8**(2): 263-271.

Marsh, P. D. (2000). "Role of the Oral Microflora in Health." Microb. Ecol. Health Dis. **12**(3).

Marsh, P. D. (2003). "Are dental diseases examples of ecological catastrophes?" Microbiology **149**(Pt 2): 279-294.

Marsh, P. D. (2003). "Plaque as a biofilm: pharmacological principles of drug delivery and action in the sub- and supragingival environment." Oral diseases **9**: 16-22.

Marsh, P. D. (2006). "Dental plaque as a biofilm and a microbial community - implications for health and disease." BMC oral health **6 Suppl 1**: S14.

Marsh, P. D., A. Featherstone, A. S. McKee, A. S. Hallsworth, C. Robinson, J. A. Weatherell, H. N. Newman and A. F. Pitter (1989). "A microbiological study of early caries of approximal surfaces in schoolchildren." J. Dent. Res. **68**(7): 1151-1154.

Marsh, P. D. and M. V. Martin (1999). Oral Microbiology. Oxford, Wright.

Marsh, P. D., A. Moter and D. A. Devine (2011). "Dental plaque biofilms: communities, conflict and control." Periodontol 2000 **55**(1): 16-35.

Marsh, P. D. and R. S. Percival (2006). "The oral microflora-friend or foe? Can we decide?" International dental journal **56**(4 Suppl 1): 233-239.

Martin, F. E., M. A. Nadkarni, N. A. Jacques and N. Hunter (2002). "Quantitative microbiological study of human carious dentine by culture and real-time PCR: association of anaerobes with histopathological changes in chronic pulpitis." J Clin Microbiol **40**(5): 1698-1704.

Matarazzo, F., A. C. Ribeiro, M. Feres, M. Faveri and M. P. Mayer (2011). "Diversity and quantitative analysis of Archaea in aggressive periodontitis and periodontally healthy subjects." J Clin Periodontol **38**(7): 621-627.

McBride, B. C. and J. S. Van der Hoeven (1981). "Role of interbacterial adherence in colonization of the oral cavities of gnotobiotic rats infected with Streptococcus mutans and Veillonella alcalescens." Infect. Immun. **33**(2): 467-472.

McConlogue, L., M. A. Brow and M. A. Innis (1988). "Structure-independent DNA amplification by PCR using 7-deaza-2'-deoxyguanosine." Nucleic Acids Res. **16**(20): 9869.

McPherson, M. J. and S. G. Moller, Eds. (2000). PCR. The Basics, BIOS Scientific Publishers Ltd.

Meckel, A. H. (1965). "The formation and properties of organic films on teeth." Arch. Oral Biol. **10**(4): 585-597.

Meng, H., L. Xu, Q. Li, J. Han and Y. Zhao (2007). "Determinants of host susceptibility in aggressive periodontitis." Periodontol 2000 **43**: 133-159.

Meurman, J. H. (2010). "Infectious and dietary risk factors of oral cancer." Oral Oncol. **46**(6): 411-413.

Miller, C. S., J. R. Berger, Y. Mootoor, S. A. Avdiushko, H. Zhu and R. J. Kryscio (2006). "High prevalence of multiple human herpesviruses in saliva from human immunodeficiency virus-infected persons in the era of highly active antiretroviral therapy." J Clin Microbiol **44**(7): 2409-2415.

Mjor, I. A. and I. Nordahl (1996). "The density and branching of dentinal tubules in human teeth." Arch. Oral Biol. **41**(5): 401-412.

Modesto, M., B. Biavati and P. Mattarelli (2006). "Occurrence of the family bifidobacteriaceae in human dental caries and plaque." Caries Res. **40**(3): 271-276.

Moran, N. A. (2006). "Symbiosis." Curr. Biol. **16**(20): R866-871.

Morge, S., E. Adamczak and L. A. Linden (1989). "Variation in human salivary pellicle formation on biomaterials during the day." Arch. Oral Biol. **34**(8): 669-674.

Moter, A. and U. B. Gobel (2000). "Fluorescence in situ hybridization (FISH) for direct visualization of microorganisms." J. Microbiol. Methods **41**(2): 85--112.

Mullis, K., F. Faloona, S. Scharf, R. Saiki, G. Horn and H. Erlich (1986). "Specific enzymatic amplification of DNA in vitro: the polymerase chain reaction." Cold Spring Harb Symp Quant Biol **51 Pt 1**: 263-273.

Munson, M. A., A. Banerjee, T. F. Watson and W. G. Wade (2004). "Molecular analysis of the microflora associated with dental caries." J Clin Microbiol **42**(7): 3023-3029.

Munson, M. A., T. Pitt-Ford, B. Chong, A. Weightman and W. G. Wade (2002). "Molecular and cultural analysis of the microflora associated with endodontic infections." J. Dent. Res. **81**(11): 761-766.

Murray, P. A., A. Prakobphol, T. Lee, C. I. Hoover and S. J. Fisher (1992). "Adherence of oral streptococci to salivary glycoproteins." Infect. Immun. **60**(1): 31-38.

Mytelka, D. S. and M. J. Chamberlin (1996). "Analysis and suppression of DNA polymerase pauses associated with a trinucleotide consensus." Nucleic Acids Res. **24**(14): 2774-2781.

Nadkarni, M. A., C. E. Caldon, K. L. Chhour, I. P. Fisher, F. E. Martin, N. A. Jacques and N. Hunter (2004). "Carious dentine provides a habitat for a complex array of novel Prevotella-like bacteria." J Clin Microbiol **42**(11): 5238-5244.

Nadkarni, M. A., M. R. Simonian, D. W. Harty, H. Zoellner, N. A. Jacques and N. Hunter (2010). "Lactobacilli are prominent in the initial stages of polymicrobial infection of dental pulp." J Clin Microbiol **48**(5): 1732-1740.

Naglik, J. R., D. L. Moyes, B. Wachtler and B. Hube (2011). "Candida albicans interactions with epithelial cells and mucosal immunity." Microbes Infect. **13**(12-13): 963-976.

Nanci, A. (2008). Ten Cate's Oral Histology: Development, Structure and Function, Mosby Elsevier.

Nemoto, K., K. Hirota, K. Murakami, K. Taniguti, H. Murata, D. Viducic and Y. Miyake (2003). "Effect of Varidase (streptodornase) on biofilm formed by Pseudomonas aeruginosa." Chemotherapy **49**(3): 121-125.

Netuschil, L., E. Reich, G. Unteregger, A. Sculean and M. Brecx (1998). "A pilot study of confocal laser scanning microscopy for the assessment of undisturbed dental plaque vitality and topography." Arch. Oral Biol. **43**(4): 277-285.

Niederman, R., T. Westernoff, C. Lee, L. L. Mark, N. Kawashima, M. Ullman-Culler, F. E. Dewhirst, B. J. Paster, D. D. Wagner, T. Mayadas, R. O. Hynes and P. Stashenko (2001). "Infection-mediated early-onset periodontal disease in P/E-selectin-deficient mice." J Clin Periodontol **28**(6): 569-575.

Noguchi, N., Y. Noiri, M. Narimatsu and S. Ebisu (2005). "Identification and localization of extraradicular biofilm-forming bacteria associated with refractory endodontic pathogens." Appl. Environ. Microbiol. **71**(12): 8738-8743.

Nyvad, B. and M. Kilian (1987). "Microbiology of the early colonization of human enamel and root surfaces in vivo." Scand J Dent Res **95**(5): 369-380.

Odaa, Y., S. Slagmana, W. G. Meijerb, L. J. Forneya and J. C. Gottschala (2000). "Influence of growth rate and starvation on fluorescent in situ hybridization of Rhodopseudomonas palustris." FEMS Microbiol. Ecol. **32**(3): 205-213.

Oligino, L. and P. Fives-Taylor (1993). "Overexpression and purification of a fimbria-associated adhesin of Streptococcus parasanguis." Infect. Immun. **61**(3): 1016--1022.

Olsen, I. and S. S. Socransky (1981). "Ultrasonic dispersion of pure cultures of plaque bacteria and plaque." Scand J Dent Res **89**(4): 307-312.

Onyido, A. E., E. S. Amadi, I. Olofin, A. A. Onwumma, I. C. Okoh and C. I. Chikwendu (2011). "Prevalence of Entamoeba gingivalis and Trichomonas tenax among dental patients attending Federal School of Dental Technology and Therapy clinic, Enugu, Nigeria." Nature and Science **9**(9): 59-62.

Orstavik, D., F. W. Kraus and L. C. Henshaw (1974). "In vitro attachment of streptococci to the tooth surface." Infect. Immun. **9**(5): 794-800.

Pace, N. R. (1997). "A molecular view of microbial diversity and the biosphere." Science **276**(5313): 734-740.

Palatinszky, M., M. Nikolausz, D. Svab and K. Marialigeti (2011). "Preferential ligation during TA-cloning of multitemplate PCR products--a factor causing bias in microbial community structure analysis." J. Microbiol. Methods **85**(2): 131-136.

Palmer, J., S. Flint and J. Brooks (2007). "Bacterial cell attachment, the beginning of a biofilm." J Ind Microbiol Biotechnol **34**(9): 577-588.

Palmer, R. J., S. M. Gordon, J. O. Cisar and P. E. Kolenbrander (2003). "Coaggregation-mediated interactions of *streptococci* and *actinomyces* detected in initial human dental plaque." J. Bacteriol. **185**(11): 3400-3409.

Palmer, R. J., K. Kazmerzak, M. C. Hansen and P. E. Kolenbrander (2001). "Mutualism versus independence: strategies of mixed-species oral biofilms in vitro using saliva as the sole nutrient source." Infect. Immun. **69**: 5794-5804.

Papke, R. T., N. B. Ramsing, M. M. Bateson and D. M. Ward (2003). "Geographical isolation in hot spring cyanobacteria." Environmental microbiology **5**(8): 650-659.

Parahitiyawa, N. B., C. Scully, W. K. Leung, W. C. Yam, L. J. Jin and L. P. Samaranayake (2010). "Exploring the oral bacterial flora: current status and future directions." Oral diseases **16**(2): 136-145.

Parisotto, T. M., C. Steiner-Oliveira, C. Duque, R. C. Peres, L. K. Rodrigues and M. Nobre-dos-Santos (2010). "Relationship among microbiological composition and presence of dental plaque, sugar exposure, social factors and different stages of early childhood caries." Arch. Oral Biol. **55**(5): 365-373.

Paster, B. J., S. K. Boches, J. L. Galvin, R. E. Ericson, C. N. Lau, V. A. Levanos, A. Sahasrabudhe and F. E. Dewhirst (2001). "Bacterial diversity in human subgingival plaque." J. Bacteriol. **183**(12): 3770-3783.

Paster, B. J., I. Olsen, J. r. A. Aas and F. E. Dewhirst (2006). "The breadth of bacterial diversity in the human periodontal pocket and other oral sites." Periodontol 2000 **42**: 80-87.

Pavlov, A. R., G. I. Belova, S. A. Kozyavkin and A. I. Slesarev (2002). "Helix-hairpin-helix motifs confer salt resistance and processivity on chimeric DNA polymerases." Proc. Natl. Acad. Sci. U. S. A. **99**(21): 13510-13515.

Pedros-Alio, C. (2012). "The rare bacterial biosphere." Annual review of marine science **4**: 449-466.

Pereni, C. I., Q. Zhao, Y. Liu and E. Abel (2006). "Surface free energy effect on bacterial retention." Colloids Surf B Biointerfaces **48**(2): 143-147.

Pernthaler, A., J. Pernthaler and R. Amann (2002). "Fluorescence in situ hybridization and catalyzed reporter deposition for the identification of marine bacteria." Appl. Environ. Microbiol. **68**(6): 3094-3101.

Petersen, F. C., L. Tao and A. A. Scheie (2005). "DNA binding-uptake system: a link between cell-to-cell communication and biofilm formation." J. Bacteriol. **187**(13): 4392-4400.

Pihlstrom, B. L., B. S. Michalowicz and N. W. Johnson (2005). "Periodontal diseases." Lancet **366**(9499): 1809-1820.

Prakobphol, A., M. J. Levine, L. A. Tabak and M. S. Reddy (1982). "Purification of a low-molecular-weight, mucin-type glycoprotein from human submandibular-sublingual saliva." Carbohydr. Res. **108**(1): 111-122.

Pratyush, D. D., S. Tiwari, A. Kumar and S. K. Singh (2012). "A new approach to touch down method using betaine as co-solvent for increased specificity and intensity of GC rich gene amplification." Gene **497**(2): 269-272.

Pruesse, E., C. Quast, K. Knittel, B. M. Fuchs, W. Ludwig, J. Peplies and F. O. Glockner (2007). "SILVA: a comprehensive online resource for quality checked and aligned ribosomal RNA sequence data compatible with ARB." Nucleic Acids Res. **35**(21): 7188-7196.

Purswani, J., A. M. Martin-Platero, P. Reboleiro-Rivas, J. Gonzalez-Lopez and C. Pozo (2011). "Comparative analysis of microbial DNA extraction protocols for groundwater samples." Anal. Biochem. **416**(2): 240-242.

Qin, Z., Y. Ou, L. Yang, Y. Zhu, T. Tolker-Nielsen, S. Molin and D. Qu (2007). "Role of autolysin-mediated DNA release in biofilm formation of Staphylococcus epidermidis." Microbiology **153**(Pt 7): 2083-2092.

Quevedo, B., E. Giertsen, V. Zijnge, H. Luthi-Schaller, B. Guggenheim, T. Thurnheer and R. Gmur (2011). "Phylogenetic group- and species-specific oligonucleotide probes for single-cell detection of lactic acid bacteria in oral biofilms." BMC Microbiol. **11**: 14.

Quince, C., A. Lanzen, T. P. Curtis, R. J. Davenport, N. Hall, I. M. Head, L. F. Read and W. T. Sloan (2009). "Accurate determination of microbial diversity from 454 pyrosequencing data." Nat Methods **6**(9): 639-641.

Ram, S., S. Kumar and M. Navazesh (2011). "Management of xerostomia and salivary gland hypofunction." Journal of the California Dental Association **39**(9): 656-659.

Ramberg, P., S. Sekino, N. G. Uzel, S. Socransky and J. Lindhe (2003). "Bacterial colonization during de novo plaque formation." J Clin Periodontol **30**(11): 990-995.

Rantakokko-Jalava, K. and J. Jalava (2002). "Optimal DNA isolation method for detection of bacteria in clinical specimens by broad-range PCR." J Clin Microbiol **40**(11): 4211-4217.

Rechenberg, D. K., T. Thurnheer and M. Zehnder (2011). "Potential systematic error in laboratory experiments on microbial leakage through filled root canals: an experimental study." International endodontic journal **44**(9): 827-835.

Reddy, M. S., L. A. Bobek, G. G. Haraszthy, A. R. Biesbrock and M. J. Levine (1992). "Structural features of the low-molecular-mass human salivary mucin." Biochem. J. **287 (Pt 2)**: 639-643.

Reeder, J. and R. Knight (2009). "The 'rare biosphere': a reality check." Nat Methods **6**(9): 636-637.

Rees, W. A., T. D. Yager, J. Korte and P. H. von Hippel (1993). "Betaine can eliminate the base pair composition dependence of DNA melting." Biochemistry **32**(1): 137-144.

Ritz, H. L. (1967). "Microbial population shifts in developing human dental plaque." Arch. Oral Biol. **12**(12): 1561-1568.

Robert, J. C. and M. Bonnaure-Mallet (1990). "[Types of interbacterial coaggregation in dental plaque]." Actual Odontostomatol (Paris)(172): 697-707.

Robertson, D. and A. J. Smith (2009). "The microbiology of the acute dental abscess." J. Med. Microbiol. **58**(Pt 2): 155-162.

Robijn, G. W., R. G. Gallego, D. J. van den Berg, H. Haas, J. P. Kamerling and J. F. Vliegenthart (1996). Structural characterization of the exopolysaccharide produced by Lactobacillus acidophilus LMG9433. Carbohydr Res. **288:** 203--218.

Robinson, C., S. Strafford, G. Rees, S. J. Brookes, J. Kirkham, R. C. Shore, P. S. Watson and S. Wood (2006). "Plaque biofilms: the effect of chemical environment on natural human plaque biofilm architecture." Arch. Oral Biol. **51**(11): 1006-1014.

Rosan, B., B. Appelbaum, E. Golub, D. Malamud and I. D. Mandel (1982). "Enhanced saliva-mediated bacterial aggregation and decreased bacterial adhesion in caries-resistant versus caries-susceptible individuals." Infect. Immun. **38**(3): 1056-1059.

Ruby, J. and M. Goldner (2007). "Nature of symbiosis in oral disease." J. Dent. Res. **86**(1): 8-11.

Rudney, J. D., M. A. Krig, E. K. Neuvar, A. H. Soberay and L. Iverson (1991). "Antimicrobial proteins in human unstimulated whole saliva in relation to each other, and to measures of health status, dental plaque accumulation and composition." Arch. Oral Biol. **36**(7): 497-506.

Ryan, F. P. (2004). "Human endogenous retroviruses in health and disease: a symbiotic perspective." Journal of the Royal Society of Medicine **97**(12): 560-565.

Sahdev, S., S. Saini, P. Tiwari, S. Saxena and K. Singh Saini (2007). "Amplification of GC-rich genes by following a combination strategy of primer design, enhancers and modified PCR cycle conditions." Mol. Cell. Probes **21**(4): 303-307.

Saiki, R., D. H. Gelfand, S. Stoffel, S. J. Scharf, R. Higuchi, G. T. Horn, K. B. Mullis and H. A. Erlich (1988). "Primer-Directed Enzymatic Amplification of DNA with a Thermostable DNA Polymerase." Science **239**(4839): 487-491.

Sakata, S., C. S. Ryu, M. Kitahara, M. Sakamoto, H. Hayashi, M. Fukuyama and Y. Benno (2006). "Characterization of the genus Bifidobacterium by automated ribotyping and 16S rRNA gene sequences." Microbiol. Immunol. **50**(1): 1-10.

Sampson, J. S., S. P. O'Connor, A. R. Stinson, J. A. Tharpe and H. Russell (1994). "Cloning and nucleotide sequence analysis of psaA, the Streptococcus pneumoniae gene encoding a 37-kilodalton protein homologous to previously reported Streptococcus sp. adhesins." Infect. Immun. **62**(1): 319-324.

Sansone, C., J. Van Houte, K. Joshipura, R. Kent and H. C. Margolis (1993). "The association of mutans streptococci and non-mutans streptococci capable of acidogenesis at a low pH with dental caries on enamel and root surfaces." J. Dent. Res. **72**(2): 508-516.

Santos, A. L., J. F. Siqueira, Jr., I. N. Rocas, E. C. Jesus, A. S. Rosado and J. M. Tiedje (2011). "Comparing the bacterial diversity of acute and chronic dental root canal infections." PLoS One **6**(11): e28088.

Sarkar, G., S. Kapelner and S. S. Sommer (1990). "Formamide can dramatically improve the specificity of PCR." Nucleic Acids Res. **18**(24): 7465.

Scannapieco, F. A. (1994). "Saliva-bacterium interactions in oral microbial ecology." Crit. Rev. Oral Biol. Med. **5**(3-4): 203-248.

Scheie, A. A. (1994). "Mechanisms of dental plaque formation." Advances in Dental Research **8**(2): 246-253.

Schlesinger, D. H. and D. I. Hay (1977). "Complete covalent structure of statherin, a tyrosine-rich acidic peptide which inhibits calcium phosphate precipitation from human parotid saliva." J. Biol. Chem. **252**(5): 1689-1695.

Schloss, P. D. and S. L. Westcott (2011). "Assessing and improving methods used in operational taxonomic unit-based approaches for 16S rRNA gene sequence analysis." Appl. Environ. Microbiol. **77**(10): 3219-3226.

Schloss, P. D., S. L. Westcott, T. Ryabin, J. R. Hall, M. Hartmann, E. B. Hollister, R. A. Lesniewski, B. B. Oakley, D. H. Parks, C. J. Robinson, J. W. Sahl, B. Stres, G. G. Thallinger, D. J. Van Horn and C. F. Weber (2009). "Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities." Appl. Environ. Microbiol. **75**(23): 7537-7541.

Schramm, A., B. M. Fuchs, J. L. Nielsen, M. Tonolla and D. A. Stahl (2002). "Fluorescence in situ hybridization of 16S rRNA gene clones (Clone-FISH) for probe validation and screening of clone libraries." Environmental microbiology **4**(11): 713-720.

Schultz-Haudt, S., M. A. Bruce and B. G. Bibby (1954). "Bacterial factors in nonspecific gingivitis." J. Dent. Res. **33**(4): 454-458.

Schwientek, P., R. Szczepanowski, C. Ruckert, J. Stoye and A. Puhler (2011). "Sequencing of high G+C microbial genomes using the ultrafast pyrosequencing technology." J. Biotechnol. **155**(1): 68-77.

Segata, N., J. Izard, L. Waldron, D. Gevers, L. Miropolsky, W. S. Garrett and C. Huttenhower (2011). "Metagenomic biomarker discovery and explanation." Genome biology **12**(6): R60.

Sekirov, I., S. L. Russell, L. C. Antunes and B. B. Finlay (2010). "Gut microbiota in health and disease." Physiol. Rev. **90**(3): 859-904.

Shaw, A. K., A. L. Halpern, K. Beeson, B. Tran, J. C. Venter and J. B. Martiny (2008). "It's all relative: ranking the diversity of aquatic bacterial communities." Environmental microbiology **10**(9): 2200-2210.

Sim, K., M. J. Cox, H. Wopereis, R. Martin, J. Knol, M. S. Li, W. O. Cookson, M. F. Moffatt and J. S. Kroll (2012). "Improved detection of bifidobacteria with optimised 16S rRNA-gene based pyrosequencing." PLoS One **7**(3): e32543.

Sipos, R., A. J. Székely, M. Palatinszky, S. Révész, K. Márialigeti and M. Nikolausz (2007). "Effect of primer mismatch, annealing temperature and PCR cycle number on 16S rRNA gene-targetting bacterial community analysis." FEMS Microbiol. Ecol. **60**(2): 341-350.

Siqueira, W. L., E. J. Helmerhorst, W. Zhang, E. Salih and F. G. Oppenheim (2007). "Acquired enamel pellicle and its potential role in oral diagnostics." Ann. N. Y. Acad. Sci. **1098**: 504-509.

Smith, A. J., B. A. Scheven, Y. Takahashi, J. L. Ferracane, R. M. Shelton and P. R. Cooper (2012). "Dentine as a bioactive extracellular matrix." Arch. Oral Biol. **57**(2): 109-121.

Soames, J. V. and J. C. Southam (1993). Oral pathology, Oxford University Press.

Socransky, S. S., R. J. Gibbons, A. C. Dale, L. Bortnick, E. Rosenthal and J. B. Macdonald (1963). "The microbiota of the gingival crevice area of man. I. Total microscopic and viable counts and counts of specific organisms." Arch. Oral Biol. **8**: 275-280.

Socransky, S. S. and A. D. Haffajee (2005). "Periodontal microbial ecology." Periodontol 2000 **38**: 135-187.

Socransky, S. S., A. D. Haffajee, M. A. Cugini, C. Smith and R. L. Kent (1998). "Microbial complexes in subgingival plaque." J Clin Periodontol **25**(2): 134-144.

Sogin, M. L., H. G. Morrison, J. A. Huber, D. Mark Welch, S. M. Huse, P. R. Neal, J. M. Arrieta and G. J. Herndl (2006). "Microbial diversity in the deep sea and the underexplored "rare biosphere"." Proc. Natl. Acad. Sci. U. S. A. **103**(32): 12115-12120.

Spiess, A. N., N. Mueller and R. Ivell (2004). "Trehalose is a potent PCR enhancer: lowering of DNA melting temperature and thermal stabilization of taq polymerase by the disaccharide trehalose." Clin. Chem. **50**(7): 1256-1259.

Stahl, D. A., D. J. Lane, G. J. Olsen and N. R. Pace (1985). "Characterization of a Yellowstone hot spring microbial community by 5S rRNA sequences." Appl. Environ. Microbiol. **49**(6): 1379-1384.

Staley, J. T. and A. Konopka (1985). "Measurement of in situ activities of nonphotosynthetic microorganisms in aquatic and terrestrial habitats." Annu. Rev. Microbiol. **39**: 321-346.

Stanley, H. R., C. L. White and L. McCray (1966). "The rate of tertiary (reparative) dentine formation in the human tooth." Oral surgery, oral medicine, and oral pathology **21**(2): 180-189.

Stephan, R. M. and E. S. Hemmens (1947). "Studies of changes in pH produced by pure cultures of oral micro-organisms; effects of varying the microbic cell concentration; comparison of different micro-organisms and different substrates; some effects of mixing certain micro-organisms." J. Dent. Res. **26**(1): 15-41.

Strevett, K. A. and G. Chen (2003). "Microbial surface thermodynamics and applications." Res. Microbiol. **154**(5): 329-335.

Su, N., C. L. Marek, V. Ching and M. Grushka (2011). "Caries prevention for patients with dry mouth." J Can Dent Assoc **77**: b85.

Suau, A., R. Bonnet, M. Sutren, J. J. Godon, G. R. Gibson, M. D. Collins and J. Dore (1999). "Direct analysis of genes encoding 16S rRNA from complex communities reveals many novel molecular species within the human gut." Appl. Environ. Microbiol. **65**(11): 4799-4807.

Suddick, R. P. and N. O. Harris (1990). "Historical perspectives of oral biology: a series." Crit. Rev. Oral Biol. Med. **1**(2): 135-151.

Sutherland, I. W. (2001). "The biofilm matrix--an immobilized but dynamic microbial environment." Trends Microbiol. **9**(5): 222-227.

Suzuki, S., T. Sreenath, N. Haruyama, C. Honeycutt, A. Terse, A. Cho, T. Kohler, R. Muller, M. Goldberg and A. B. Kulkarni (2009). "Dentin sialoprotein and dentin phosphoprotein have distinct roles in dentin mineralization." Matrix Biol. **28**(4): 221-229.

Svensater, G., M. Borgstrom, G. H. Bowden and S. Edwardsson (2003). "The acid-tolerant microbiota associated with plaque from initial caries and healthy tooth surfaces." Caries Res. **37**(6): 395-403.

Svensäter, G., J. Welin, J. C. Wilkins, D. Beighton and I. R. Hamilton (2001). "Protein expression by planktonic and biofilm cells of Streptococcus mutans." FEMS Microbiol. Lett. **205**(1): 139--146.

Takahashi, N. and B. Nyvad (2008). "Caries ecology revisited: microbial dynamics and the caries process." Caries Res. **42**(6): 409-418.

Takahashi, N. and B. Nyvad (2011). "The role of bacteria in the caries process: ecological perspectives." J. Dent. Res. **90**(3): 294-303.

Takahashi, N. and T. Yamada (1999). "Acid-induced acid tolerance and acidogenicity of non-mutans streptococci." Oral Microbiol. Immunol. **14**(1): 43-48.

Tanner, A., M. F. Maiden, P. J. Macuch, L. L. Murray and R. L. Kent, Jr. (1998). "Microbiota of health, gingivitis, and initial periodontitis." J Clin Periodontol **25**(2): 85-98.

Tanner, A. C., J. M. Mathney, R. L. Kent, N. I. Chalmers, C. V. Hughes, C. Y. Loo, N. Pradhan, E. Kanasi, J. Hwang, M. A. Dahlan, E. Papadopolou and F. E. Dewhirst (2011). "Cultivable anaerobic microbiota of severe early childhood caries." J Clin Microbiol **49**(4): 1464-1474.

Taylor, D. L., I. C. Herriott, J. Long and K. O'Neill (2007). "TOPO TA is A-OK: a test of phylogenetic bias in fungal environmental clone library construction." Environmental microbiology **9**(5): 1329-1334.

Theilade, E. (1986). "The non-specific theory in microbial etiology of inflammatory periodontal diseases." J Clin Periodontol **13**(10): 905-911.

Theilade, E. (1986). "The nonspecific theory in microbial etiology of inlammatory periodontal-diseases." Journal of Clinical Periodontology **13**(10): 905-911.

Theilade, E., W. H. Wright, S. B. Jensen and H. Loe (1966). "Experimental gingivitis in man II." Journal of Periodontal Research **1**: 1-13.

Theilade, E., W. H. Wright, S. B. Jensen and H. Loe (1966). "Experimental gingivitis in man. II. A longitudinal clinical and bacteriological investigation." J. Periodontal Res. **1**: 1-13.

Thurnheer, T., R. Gmür, E. Giertsen and B. Guggenheim (2001). "Automated fluorescent in situ hybridization for the specific detection and quantification of oral streptococci in dental plaque." J. Microbiol. Methods **44**(1): 39-47.

Thurnheer, T., R. Gmür and B. Guggenheim (2004). "Multiplex FISH analysis of a six-species bacterial biofilm." J. Microbiol. Methods **56**(1): 37-47.

Torlakovic, L., V. Klepac-Ceraj, B. Ogaard, S. L. Cotton, B. J. Paster and I. Olsen (2012). "Microbial community succession on developing lesions on human enamel." Journal of oral microbiology **4**.

Urwin, R. and M. C. Maiden (2003). "Multi-locus sequence typing: a tool for global epidemiology." Trends Microbiol. **11**(10): 479-487.

Vadillo-Rodríguez, V., H. J. Busscher, H. C. van der Mei, J. de Vries and W. Norde (2005). "Role of lactobacillus cell surface hydrophobicity as probed by AFM in adhesion to surfaces at low and high ionic strength." Colloids Surf B Biointerfaces **41**(1): 33-41.

van der Mei, H. C., M. Rustema-Abbing, J. de Vries and H. J. Busscher (2008). "Bond strengthening in oral bacterial adhesion to salivary conditioning films." Appl. Environ. Microbiol. **74**(17): 5511-5515.

van Houte, J., J. Lopman and R. Kent (1996). "The final pH of bacteria comprising the predominant flora on sound and carious human root and enamel surfaces." J. Dent. Res. **75**(4): 1008-1014.

van Loosdrecht, M. C., J. Lyklema, W. Norde, G. Schraa and A. J. Zehnder (1987). "Electrophoretic mobility and hydrophobicity as a measured to predict the initial steps of bacterial adhesion." Appl. Environ. Microbiol. **53**(8): 1898-1901.

van Ruyven, F. O., P. Lingstrom, J. van Houte and R. Kent (2000). "Relationship among mutans streptococci, "low-pH" bacteria, and lodophilic polysaccharide-

producing bacteria in dental plaque and early enamel caries in humans." J. Dent. Res. **79**(2): 778-784.

van Strijp, A. J., T. J. van Steenbergen, J. de Graaff and J. M. ten Cate (1994). "Bacterial colonization and degradation of demineralized dentin matrix in situ." Caries Res. **28**(1): 21-27.

Varadaraj, K. and D. M. Skinner (1994). "Denaturants or cosolvents improve the specificity of PCR amplification of a G + C-rich DNA using genetically engineered DNA polymerases." Gene **140**(1): 1-5.

Vartoukian, S. R., R. M. Palmer and W. G. Wade (2009). "Diversity and morphology of members of the phylum "synergistetes" in periodontal health and disease." Appl. Environ. Microbiol. **75**(11): 3777-3786.

Vartoukian, S. R., R. M. Palmer and W. G. Wade (2010). "Strategies for culture of 'unculturable' bacteria." FEMS Microbiol. Lett. **309**(1): 1-7.

Veerman, E. C., A. J. Ligtenberg, L. C. Schenkels, E. Walgreen-Weterings and A. V. Nieuw Amerongen (1995). "Binding of human high-molecular-weight salivary mucins (MG1) to Hemophilus parainfluenzae." J. Dent. Res. **74**(1): 351-357.

Wade, W. (2002). "Unculturable bacteria--the uncharacterized organisms that cause oral infections." Journal of the Royal Society of Medicine **95**(2): 81-83.

Wade, W. G. (2011). "Has the use of molecular methods for the characterization of the human oral microbiome changed our understanding of the role of bacteria in the pathogenesis of periodontal disease?" J Clin Periodontol **38 Suppl 11**: 7-16.

Wagner, A., N. Blackstone, P. Cartwright, M. Dick, B. Misof, P. Snow, G. P. Wagner, J. Bartels, M. Murtha and J. Pendleton (1994). "Surveys of Gene Families Using Polymerase Chain Reaction: PCR Selection and PCR Drift." Syst Biol **43**(2): 250-261.

Wang, Y. and P. Y. Qian (2009). "Conservative fragments in bacterial 16S rRNA genes and primer design for 16S ribosomal DNA amplicons in metagenomic studies." PLoS One **4**(10): e7401.

Ward, D. M., R. Weller and M. M. Bateson (1990). "16S rRNA sequences reveal numerous uncultured microorganisms in a natural community." Nature **345**(6270): 63-65.

Weisburg, W. G., S. M. Barns, D. A. Pelletier and D. J. Lane (1991). "16S ribosomal DNA amplification for phylogenetic study." J. Bacteriol. **173**(2): 697-703.

Weiss, E. I., J. London, P. E. Kolenbrander, A. R. Hand and R. Siraganian (1988). "Localization and enumeration of fimbria-associated adhesins of Becteroides loescheii." Journal of Bacteriology **170f**: 1123-1128.

Weitere, M., T. Bergfeld, S. A. Rice, C. Matz and S. Kjelleberg (2005). "Grazing resistance of Pseudomonas aeruginosa biofilms depends on type of protective mechanism, developmental stage and protozoan feeding mode." Environmental microbiology **7**(10): 1593-1601.

Welin, J., J. C. Wilkins, D. Beighton and G. Svensäter (2004). "Protein expression by Streptococcus mutans during initial stage of biofilm formation." Appl. Environ. Microbiol. **70**(6): 3736--3741.

Whitchurch, C. B., T. Tolker-Nielsen, P. C. Ragas and J. S. Mattick (2002). "Extracellular DNA required for bacterial biofilm formation." Science **295**(5559): 1487.

Whitford, M. F., R. J. Forster, C. E. Beard, J. Gong and R. M. Teather (1998). "Phylogenetic analysis of rumen bacteria by comparative sequence analysis of cloned 16S rRNA genes." Anaerobe **4**(3): 153-163.

Whittaker, C. J., C. M. Klier and P. E. Kolenbrander (1996). "Mechanisms of adhesion by oral bacteria." Annu. Rev. Microbiol. **50**: 513-552.

Whittaker, C. J., C. M. Klier and P. E. Kolenbrander (1996). Mechanisms of adhesion by oral bacteria. Annu Rev Microbiol. **50:** 513--552.

Wilson, K. H., R. B. Blitchington and R. C. Greene (1990). "Amplification of bacterial 16S ribosomal DNA with polymerase chain reaction." J Clin Microbiol **28**(9): 1942-1946.

Wimpenny, J. W. T. and R. Colasanti (1997). "A unifying hypothesis for the structure of microbial biofilms based on cellular automaton models." FEMS Microbiol. Ecol. **22**: 1-16.

Woese, C. R. (1987). "Bacterial evolution." Microbiol Rev **51**(2): 221-271.

Woese, C. R. and G. E. Fox (1977). "Phylogenetic structure of the prokaryotic domain: the primary kingdoms." Proc. Natl. Acad. Sci. U. S. A. **74**(11): 5088-5090.

Wolfaardt, Lawrence, Robarts and Caldwell (1998). "In situ Characterization of Biofilm Exopolymers Involved in the Accumulation of Chlorinated Organics." Microb. Ecol. **35**(3): 213-223.

Wood, S. R., J. Kirkham, P. D. Marsh, R. C. Shore, B. Nattress and C. Robinson (2000). "Architecture of intact natural human plaque biofilms studied by confocal laser scanning microscopy." J. Dent. Res. **79**(1): 21-27.

Wu, J. H., P. Y. Hong and W. T. Liu (2009). "Quantitative effects of position and type of single mismatch on single base primer extension." J. Microbiol. Methods **77**(3): 267-275.

Wyss, C. (1989). "Dependence of proliferation of Bacteroides forsythus on exogenous N-acetylmuramic acid." Infect. Immun. **57**(6): 1757-1759.

Xiao, J., M. I. Klein, M. L. Falsetta, B. Lu, C. M. Delahunty, J. R. Yates, 3rd, A. Heydorn and H. Koo (2012). "The Exopolysaccharide Matrix Modulates the Interaction between 3D Architecture and Virulence of a Mixed-Species Oral Biofilm." PLoS pathogens **8**(4): e1002623.

Xie, Q., K. Hong and M. Goodfellow (2011). "Genus-specific primers targeting the 16S rRNA gene for PCR detection of members of the genus Verrucosispora." Antonie Leeuwenhoek **100**(1): 117-128.

Yang, F., X. Zeng, K. Ning, K. L. Liu, C. C. Lo, W. Wang, J. Chen, D. Wang, R. Huang, X. Chang, P. S. Chain, G. Xie, J. Ling and J. Xu (2012). "Saliva microbiomes distinguish caries-active from healthy human populations." The ISME journal **6**(1): 1-10.

Yao, Y., J. Grogan, M. Zehnder, U. Lendenmann, B. Nam, Z. Wu, C. E. Costello and F. G. Oppenheim (2001). "Compositional analysis of human acquired enamel pellicle by mass spectrometry." Arch. Oral Biol. **46**(4): 293-303.

Zambon, J. J. (1985). "Actinobacillus-actinomycetemcomitans in human periodontal-disease." Journal of Clinical Periodontology **12**(1): 1-20.

Zambon, J. J., J. Slots and R. J. Genco (1983). "Serology of oral *Actinobacillus-actinomycetemcomitans* and serotype distribution in human periodontal-disease." Infection and Immunity **41**(1): 19-27.

Zaura, E., B. J. Keijser, S. M. Huse and W. Crielaard (2009). "Defining the healthy "core microbiome" of oral microbial communities." BMC Microbiol. **9**: 259.

Zijnge, V., M. B. van Leeuwen, J. E. Degener, F. Abbas, T. Thurnheer, R. Gmur and H. J. Harmsen (2010). "Oral biofilm architecture on natural teeth." PLoS One **5**(2): e9321.

zur Hausen, H. (1996). "Papillomavirus infections--a major cause of human cancers." Biochim. Biophys. Acta **1288**(2): F55-78.

Appendix 1

Molecular characterisation of the bacterial community in dentinal caries - Kathrin Schulze-Schweifing - A thesis submitted in accordance with the requirements of the degree of Doctor of Philosophy for King's College London - August 2012 - Department of Clinical and Diagnostic Sciences, Dental Institute, King's College London

Table 1: Distribution of detected taxa for each patient sample (A – F) and library (I and 1- 5) using a PCR / cloning / Sanger sequencing approach. Primers used: Isolates - 27F CM / 1492R, Library 1 – 27F CM / 1492R, Library 2 – 27F YM / 1492R, Library 3 – 39F / 1387R, Library 4 - 39F / 1492R, Library 5 – 61F / 1387R

	A I	A 1	A 2	A 3	A 4	A 5	B I	B 1	B 2	B 3	B 4	B 5	C I	C 1	C 2	C 3	C 4	C 5	D I	D 1	D 2	D 3	D 4	D 5	E I	E 1	E 2	E 3	E 4	E 5	F I	F 1	F 2	F 3	F 4	F 5
	9 0	9 0	9 0	9 0	9 0	9 0	9 0	9 0	9 0	9 0	9 0	9 0	9 0	9 0	9 0	9 0	9 0	9 0	9 0	9 0	9 0	9 0	9 0	9 0	9 0	9 0	9 0	9 0	9 0	9 0	9 0	9 0	9 0	9 0	9 0	9 0
***Abiotrophia defectiva* OT 389**	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	1	0
***Anaeroglobus geminatus* OT 121**	0	0	0	0	0	0	1	1	0	0	3	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
***Flavabacteriaceae* K1**	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
***Bergeyella* sp. OT 322**	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
***Catonella morbi* OT 165**	0	0	0	0	0	0	0	3	2	0	0	2	0	2	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0

***Catonella* sp. OT 164**	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
***Centipeda periodontii* OT 726**	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	7	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Clostridiales[F-1][G-1] sp. OT 093	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
***Clostridiales*[F-2][G-1] sp. OT 075**	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
***Clostridiales*[F-2][G-2] sp. OT 085**	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
***Dialister invisus* OT 118**	0	3	0	0	0	0	0	0	0	0	1	0	0	1	0	3	3	0	2	5	4	2	6	4	0	0	0	0	0	0	0	0	0	0	0	0
***Dialister pneumosintes* OT 736**	0	0	0	0	0	0	0	1	0	0	0	0	1	0	2	0	5	1	1	1	0	1	1	1	0	0	0	0	0	1	0	0	0	0	0	0
***Eikenella corrodens* OT 577**	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0
***Erysipelotrichaceae* K1**	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
***Erysipelotrichaceae* K2**	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
***Eubacterium*[11][G-1] *infirmum* OT 105**	0	0	0	0	0	0	3	0	0	0	1	0	2	2	0	0	0	4	8	10	12	1	3	2	0	0	0	0	0	0	0	0	0	0	0	0

| | | |

***Lachnospiraceae*[G-2] sp. OT 100**	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
***Lachnospiraceae*[G-4] sp. OT 419**	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	2	1	0
Lactobacillus casei	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	1	1	0	0	0	0	0	0
***Lactobacillus crispatus* OT 817**	0	1	1	1	9	1	7	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
***Lactobacillus fermentum* OT 608**	1	1	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	3	3	5	2	6	0	0	0	0	0
***Lactobacillus gasseri* OT 615**	5	5	5	4	6	3	8	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	4	4	3	3	2	0	0	2	1	1	0
	3	2	6	9	0																							5	8	6	8	5	3					
Lactobacillus helveticus	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	2	1	0	0	0	0	0
Lactobacillus mucosae	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Lactobacillus nagelii	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	3	1	0	0
***Lactobacillus oris* OT 709**	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	0	0	1	0	1	0	0	0	0	0
***Lactobacillus paracasei* OT 716**	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	3	1	1	1	5	0	0	0	0	0	0

Lactobacillus rapi	0	2	1	7	4	5	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Lactobacillus rhamnosus **OT 749**	14	2	3	2	2	5	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	17	7	1	5	7	3	0	0	0	0	0	0
Lactobacillus **sp. OT 418**	0	5	2	5	4	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Lactobacillus vaginalis **OT 051**	20	9	11	6	8	9	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	0	0	2	4	3	0	0	0	0	0	0
Megasphaera **sp. OT 123**	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Megasphera micronucliformis **OT 122**	0	0	0	0	0	0	0	0	0	0	0	2	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	3	0	1	1
Mogibacterium diversum / neglectum / vescum	0	0	0	0	0	0	0	1	0	0	1	0	0	0	0	0	0	0	0	3	1	0	0	2	1	0	0	0	0	0	0	0	0	4	0	0	0
Mogibacterium timidum **OT 042**	0	0	0	0	0	0	1	1	0	0	0	0	2	0	0	0	0	0	1	5	2	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Oribacterium sinus **OT 457**	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Oribacterium **sp. OT 102**	0	0	0	0	0	0	1	0	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Oribacterium **sp. OT 372 / OT**	0	0	0	0	0	0	0	0	0	0	0	0	0	1	6	0	1	0	0	1	2	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0	0

078																																							
***Parvimonas micros* OT 111**	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	6	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
***Parvimonas* sp. OT 393**	0	0	0	0	0	0	0	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
***Peptococcus* sp. OT 167**	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
***Peptostreptococcaceae[11]*[G -2] sp. OT 091**	0	0	0	0	0	0	1	0	0	0	0	2	0	0	0	0	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
***Peptostreptococcus stomatis* OT 112**	0	0	0	0	0	0	3	1	0	0	0	0	0	2	7	0	3	2	0	0	0	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
***Pseudoramibacter alactolyticus* OT 538**	0	0	0	0	0	0	0	1	0	0	3	0	1 6	1 8	1 9	8	2 2	5	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
***Selenomonas dianae* OT 139**	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
***Selenomonas infelix* OT 639**	0	0	0	0	0	0	1	1	1	0	1	0	0	1	0	0	0	0	0	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
***Selenomonas noxia* OT 130**	0	0	0	0	0	0	2	2	0	0	0	2	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0
***Selenomonas* sp. OT 126**	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
***Selenomonas* sp. OT 136**	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0	0	0	0	0

Selenomonas* sp. OT 137 / *artemidis	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0
***Selenomonas* sp. OT 138**	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
***Selenomonas* sp. OT 138 / *infelix* / OT 146**	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
***Selenomonas* sp. OT 139**	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
***Selenomonas* sp. OT 149**	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0
***Selenomonas* sp. OT 478**	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0
***Selenomonas* sp. OT 479**	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	1	0	0	0	0	0	1	0	0	0	0	0	0	0	0	1	0	0	0	0
***Selenomonas sputigena* OT 151**	0	0	0	0	0	0	3	0	1	0	1	0	0	0	0	0	0	0	2	2	1	3	1	1	0	0	0	0	0	0	0	0	0	1	0	0
***Shuttleworthia satelles* OT 095**	0	0	0	0	0	0	0	0	0	0	1	0	6	3	2	0	0	2	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
***Solobacterium moorei* OT 678**	0	0	0	0	0	0	2	1	0	1	0	0	0	1	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
***Sphaerocytophaga* S3 sp. OT 337**	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	1	1	4

| | | |

| | | |

| | | |

***Campylobacter gracilis* OT 623**	0	0	0	0	0	0	2	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0
***Campylobacter showae* OT 763**	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	3	0	1	0
***Cardiobacterium hominis* OT 633**	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	0	0	0	0	0	0	0	1	1	0	0	0
***Haemophilus* K1**	0	1	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
***Haemophilus parainfluenzae* OT 718**	2	0	0	0	0	0	0	2	2	1	1	6	0	0	0	0	0	0	0	0	0	0	0	0	0	4	3	2	2	9	1	1	0	0	2	3
***Haemophilus* sp. OT 035**	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0
***Haemophilus* sp. OT 036**	0	0	0	0	0	3	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1	1	0	0	0	0	0	0
***Haemophilus* sp. OT C82**	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0
***Kingella oralis* OT 706**	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	1	0	0	0	2	0
***Comamonadaceae* K1**	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
***Lautropia mirabilis* OT 022**	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0
***Neisseria bacilliformis* sp. OT**	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	5	0

013																																					
***Neisseria elongata* OT 598**	0	0	0	0	0	0	0	1	0	1	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	1	1	2	0
Neisseria flava / mucosa / pharyngis / sicca	0	0	0	0	0	0	1	4	2	3	3	0	0	0	0	0	0	0	0	2	0	0	3	0	0	0	1	0	0	0	0	1	2	0	3	0
Neisseria flavescens / subflava	0	0	0	0	0	0	0	2	2	2	6	1	0	0	0	0	0	0	0	0	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0
***Proteus mirabolis* OT 676**	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0
***Actinomyces* K1**	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0
***Actinomyces odontolyticus* OT 701**	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Actinomyces oris	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	1	1	1	1	0	0	0	0	0	0
***Actinomyces* sp. OT 169**	0	0	0	2	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
***Actinomyces* sp. OT 170**	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
***Actinomyces* sp. OT 171**	0	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
***Actinomyces* sp. OT 180**	0	0	0	0	0	0	0	0	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

***Actinomyces* sp. OT 449**	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1	0	1
***Atopobium parvulum* OT 723**	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	2	0	1	0	0	0	0	0	0	0	0	0	0	1	0	0
***Atopobium rimae* OT 750**	0	0	0	0	0	0	1	0	0	0	0	0	16	2	1	0	3	2	6	2	1	1	0	1	0	0	0	0	0	0	0	0	0	0	0
***Atopobium* sp. OT 416**	0	0	0	0	0	0	0	0	0	0	0	0	10	5	3	5	2	23	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
***Bifidobacterium dentium* OT 588**	0	0	0	0	0	0	0	0	0	0	0	0	7	0	4	8	1	7	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
***Parascardovia denticolens* OT 586**	0	0	0	0	0	0	8	0	1	0	0	0	0	0	1	0	0	0	6	0	0	2	0	2	2	0	0	4	1	3	0	0	3	2	2	6
***Scardovia inopinata* OT 642**	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	1	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
***Cryptobacterium curtum* OT 579**	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0
***Olsenella profusa* OT 806**	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	19	0	0	0	0	0	1	0	0	0	0	0	37	0	0	0	0	0
***Olsenella uli* OT 038**	0	0	0	0	0	0	7	0	0	0	0	0	6	1	0	0	1	0	6	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
***Propionibacterium acnes* OT**	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

530

***Propionibacterium* sp. OT 191**	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	0	0	0	1	0	1 5	0	1	0	0	0
***Rothia dentocariosa* OT 587**	0	0	1	1	1	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	5	0	1	0	0	0	0
***Rothia mucilaginosa* OT 681**	0	0	0	0	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0
***Slackia exigua* OT 602**	0	0	0	0	0	0	0	0	0	0	0	0	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
***Bacteroidetes* K1**	0	0	0	0	0	0	0	0	2	1	0	4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
***Bacteroidetes*[G-1] sp. OT 272**	0	0	0	0	0	0	5	1	3	4	1	2	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
***Bacteroidetes*[G-2] sp. OT 274**	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0
***Bacteroidetes*[G-5] sp. OT 511**	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	0	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Capnocytophaga gingivalis / granulosa	0	0	0	0	0	0	1	1	3	0	4	0	0	0	0	2	0	0	0	0	0	0	1	0	0	0	0	0	0	0	1	0	0	1	1	0	2
***Capnocytophaga* K2**	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	1	0	0	1	0

***Capnocytophaga* K1**	0	0	0	0	0	0	0	0	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
***Capnocytophaga* sp. OT 326**	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0
***Capnocytophaga* sp. OT 329**	0	0	0	0	0	0	0	0	0	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	3	0	1	3	1	4
***Capnocytophaga* sp. OT 332**	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
***Capnocytophaga* sp. OT 335**	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	1
***Capnocytophaga* sp. OT 336**	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	3
***Capnocytophaga sputigena* OT 775**	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	1	0	0	3	0	0	1
***Porphyromonas catoniae* OT 283**	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	1	0	0	0
***Porphyromonas endodontalis* OT 273**	0	0	0	0	0	0	0	2	2	7	8	0	0	0	0	2	0	0	0	1	2	2	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0
***Porphyromonas gingivalis* OT 619**	0	0	0	0	0	0	4	6	4	8	5	5	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
***Porphyromonas* sp. OT 278**	0	0	0	0	0	0	0	0	0	1	1	2	0	0	0	0	0	0	0	3	2	2	0	6	0	0	0	0	0	0	0	0	0	0	0	0	0
***Porphyromonas* sp. OT 279**	0	0	0	0	0	1	0	0	1	2	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	1	

***Prevotella baroniae* OT 553**	0	0	0	0	0	0	0	0	0	0	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
***Prevotella buccae* OT 560**	0	0	0	0	0	0	3	0	1	2	1	2	1	0	0	0	1	0	3	2	1	5	5	4	0	0	0	0	0	0	0	0	0	0	0	0
***Prevotella dentalis* OT 583**	0	0	0	0	0	0	1	0	0	1	0	4	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
***Prevotella denticola* OT 291**	0	0	0	0	0	0	1	1	0	1	1	3	9	7	1 0	1 6	8	1 2	4	4	4	1	5	1 0	0	0	0	0	0	0	0	1 5	9	5	1 0	1 3	2 4
***Prevotella enoeca* OT 600**	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
***Prevotella intermedia* OT 643**	0	0	0	0	0	0	1	1	0	0	0	3	4	7	3	3	9	8	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
***Prevotella* K1**	0	0	0	0	0	0	0	1	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
***Prevotella* K3**	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	2	0	1	0	0	0	0	0	0	0	0	0	0	0	0
***Prevotella* K2**	0	0	0	0	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Prevotella maculosa	0	0	0	0	0	0	3	1	1	1	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
***Prevotella marshii* OT 665**	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	5	1	0	0	0	0	0	0	0	0	0	0	0	0
***Prevotella melaninogenica* OT 469**	0	0	0	0	0	0	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	1	2	0	4	2	1	3	1 1
***Prevotella micans* OT 378**	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0

***Prevotella multisaccarivorax* OT 794**	0	0	0	0	0	0	0	0	0	0	0	2	0	7	1 5	1 2	5	1 1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
***Prevotella nigrescens* OT 693**	0	1	0	0	0	0	1	2	2	0	0	2	1	0	3	3	2	2	2	4	3	4	3	6	0	0	0	0	0	0	0	0	0	1	0	0			
***Prevotella oralis* OT 705**	0	0	0	0	0	0	3	2	2	8	3	2	0	0	0	1	0	0	3	2	3	6	0	1	0	0	0	0	0	0	0	0	0	0	0	0			
***Prevotella oris* OT 311**	0	0	0	0	0	0	0	1	1	0	0	0	0	0	1	1	2	1	0	1	1	0	0	2	0	0	0	1	0	0	0	0	0	0	0	1			
***Prevotella oulora* OT 288**	0	0	0	0	0	0	0	0	0	1	0	0	1	0	0	0	1	0	0	0	1	0	2	0	0	0	0	0	0	0	0	0	1	0	0	0			
***Prevotella pallens* OT 714**	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	0	0	0	0	0	0	0	2	0	0	0	0			
***Prevotella salivae* OT 307**	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	1	1			
***Prevotella* sp. OT 292**	0	0	0	0	0	0	0	1	1	1	1	1	0	0	0	0	3	0	0	0	0	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0			
***Prevotella* sp. OT 300**	0	0	0	0	0	0	0	0	1	1	0	0	0	1	1	2	0	0	0	1	0	3	0	3	0	0	0	0	0	0	0	0	1	0	0	0			
***Prevotella* sp. OT 302**	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	3	3	0	1	1	0	1	5	0	0	0	0	0	0	0	0	0	0	0	0			
***Prevotella* sp. OT 303 / OT 296**	0	0	0	0	0	0	0	0	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0			
***Prevotella* sp. OT 308**	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	1	0	0	0	0	0	0			
***Prevotella* sp. OT 314 / OT**	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0			

| | | |

Tannerella forsythia	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
***Porphyromonodaceae* K1**	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0
***Tannerella* sp. OT 286**	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	1
***Fusobacterium nucleatum* ss *animalis* OT 420**	0	0	0	0	0	0	1	3	2	0	6	0	0	0	0	0	0	0	0	1	4	1	1	1	0	0	0	0	0	0	1	1	1	0	1	0
***Fusobacterium nucleatum* ss *polymorphum* OT 202**	0	0	0	0	0	0	0	1	2	2	1	0	0	0	0	0	0	0	0	0	2	0	0	0	0	0	0	0	0	0	0	2	1	2	4	2
***Fusobacterium nucleatum* ss *vincentii* OT 200**	0	0	0	0	0	0	0	0	1	1	1	0	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0
***Fusobacterium periodontium* OT 201**	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0
***Leptotrichia* sp. OT 212**	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	1	0
***Leptotrichia* sp. OT 213**	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	
***Leptotrichia* sp. OT 215**	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	3	1	1	0	1	
***Leptotrichia* sp. OT 392**	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	
***Leptotrichia* sp. OT 417 / OT**	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	

| | | |

***Treponema* sp. OT 231**	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
***Treponema* sp. OT 262**	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
***Treponema* sp. OT 268**	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0
***Treponema* sp. OT 270**	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0
***Treponema vincentii* OT 029**	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0
***Synergistes*[G-3] sp. OT 363**	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0

Appendix 2

Molecular characterisation of the bacterial community in dentinal caries - Kathrin Schulze-Schweifing - A thesis submitted in accordance with the requirements of the degree of Doctor of Philosophy for King's College London - August 2012 - Department of Clinical and Diagnostic Sciences, Dental Institute, King's College London

Table 1: 454 pyrosequencing results using primer pair 27F YM / 519R and the mothur suite to calculate taxonomic assignment, showing the distribution of detected taxa at the various taxon levels for each patient sample (A – F).

taxlevel	rankID	taxon	daughterlevels	total	KA	KB	KC	KD	KE	KF
0	0	Root	1	25758	4293	4293	4293	4293	4293	4293
1	0.1	Bacteria	11	25758	4293	4293	4293	4293	4293	4293
2	0.1.1	*Actinobacteria*	1	1638	61	318	310	417	101	431
3	0.1.1.1	*Actinobacteria*	3	1638	61	318	310	417	101	431
4	0.1.1.1.1	*Actinomycetales*	4	1031	61	286	41	189	91	363
5	0.1.1.1.1.1	*Actinomycetaceae*	1	517	26	247	28	42	25	149
6	0.1.1.1.1.1.2	*Actinomyces*	15	517	26	247	28	42	25	149
7	0.1.1.1.1.1.2.3	*israelii*	0	61	0	53	0	0	1	7
7	0.1.1.1.1.1.2.5	*naeslundii*	0	7	0	5	0	0	0	2
7	0.1.1.1.1.1.2.10	sp._oral_taxon_169	0	7	3	2	2	0	0	0
7	0.1.1.1.1.1.2.12	sp._oral_taxon_171	0	16	0	1	3	10	0	2

7	0.1.1.1.1.1.2.14	sp._oral_taxon_175	0	20	7	6	6	0	1	0
7	0.1.1.1.1.1.2.15	sp._oral_taxon_177	0	13	0	10	2	1	0	0
7	0.1.1.1.1.1.2.17	sp._oral_taxon_179	0	12	0	12	0	0	0	0
7	0.1.1.1.1.1.2.18	sp._oral_taxon_180	0	23	0	12	7	4	0	0
7	0.1.1.1.1.1.2.19	sp._oral_taxon_181	0	1	0	0	0	0	0	1
7	0.1.1.1.1.1.2.20	sp._oral_taxon_414	0	10	0	0	0	0	0	10
7	0.1.1.1.1.1.2.21	sp._oral_taxon_446	0	16	0	0	0	1	0	15
7	0.1.1.1.1.1.2.22	sp._oral_taxon_448	0	10	8	1	0	0	1	0
7	0.1.1.1.1.1.2.23	sp._oral_taxon_449	0	114	0	38	1	5	0	70
7	0.1.1.1.1.1.2.24	sp._oral_taxon_525	0	12	0	10	0	0	0	2
7	0.1.1.1.1.1.2.25	unclassified	0	195	8	97	7	21	22	40
5	0.1.1.1.1.2	*Corynebacteriaceae*	1	100	2	21	1	46	0	30
6	0.1.1.1.1.2.1	*Corynebacterium*	2	100	2	21	1	46	0	30
7	0.1.1.1.1.2.1.2	*durum*	0	1	0	0	0	1	0	0
7	0.1.1.1.1.2.1.3	*matruchotii*	0	99	2	21	1	45	0	30
5	0.1.1.1.1.5	*Micrococcaceae*	1	113	33	5	12	11	29	23
6	0.1.1.1.1.5.3	*Rothia*	3	113	33	5	12	11	29	23
7	0.1.1.1.1.5.3.1	*aeria*	0	7	1	1	0	4	1	0

7	0.1.1.1.1.5.3.2	*dentocariosa*	0	94	30	4	12	0	25	23
7	0.1.1.1.1.5.3.3	*mucilaginosa*	0	12	2	0	0	7	3	0
5	0.1.1.1.1.7	*Propionibacteriaceae*	1	301	0	13	0	90	37	161
6	0.1.1.1.1.7.1	*Propionibacterium*	2	301	0	13	0	90	37	161
7	0.1.1.1.1.7.1.1	*acidifaciens*	0	300	0	12	0	90	37	161
7	0.1.1.1.1.7.1.4	*propionicum*	0	1	0	1	0	0	0	0
4	0.1.1.1.2	*Bifidobacteriales*	1	131	0	18	33	41	10	29
5	0.1.1.1.2.1	*Bifidobacteriaceae*	3	131	0	18	33	41	10	29
6	0.1.1.1.2.1.3	*Bifidobacterium*	1	34	0	0	33	0	0	1
7	0.1.1.1.2.1.3.1	*dentium*	0	34	0	0	33	0	0	1
6	0.1.1.1.2.1.5	*Parascardovia*	1	63	0	11	0	14	10	28
7	0.1.1.1.2.1.5.1	*denticolens*	0	63	0	11	0	14	10	28
6	0.1.1.1.2.1.6	*Scardovia*	2	34	0	7	0	27	0	0
7	0.1.1.1.2.1.6.1	*inopinata*	0	33	0	7	0	26	0	0
7	0.1.1.1.2.1.6.3	*wiggsiae*	0	1	0	0	0	1	0	0
4	0.1.1.1.3	*Coriobacteriales*	1	476	0	14	236	187	0	39
5	0.1.1.1.3.1	*Coriobacteriaceae*	4	476	0	14	236	187	0	39
6	0.1.1.1.3.1.1	*Atopobium*	3	429	0	13	228	182	0	6

7	0.1.1.1.3.1.1.2	*parvulum*	0	6	0	0	0	2	0	4
7	0.1.1.1.3.1.1.3	*rimae*	0	356	0	13	161	180	0	2
7	0.1.1.1.3.1.1.6	sp._oral_taxon_416	0	67	0	0	67	0	0	0
6	0.1.1.1.3.1.2	*Cryptobacterium*	1	6	0	0	2	0	0	4
7	0.1.1.1.3.1.2.1	*curtum*	0	6	0	0	2	0	0	4
6	0.1.1.1.3.1.4	*Olsenella*	3	34	0	1	0	4	0	29
7	0.1.1.1.3.1.4.1	*profusa*	0	31	0	0	0	3	0	28
7	0.1.1.1.3.1.4.2	sp._oral_taxon_807	0	2	0	1	0	1	0	0
7	0.1.1.1.3.1.4.5	unclassified	0	1	0	0	0	0	0	1
6	0.1.1.1.3.1.5	*Slackia*	1	7	0	0	6	1	0	0
7	0.1.1.1.3.1.5.1	*exigua*	0	7	0	0	6	1	0	0
2	0.1.2	*Bacteroidetes*	2	5210	3	935	1898	1779	20	575
3	0.1.2.1	*Bacteroides*	1	5029	3	854	1894	1732	20	526
4	0.1.2.1.1	*Bacteroidales*	7	5029	3	854	1894	1732	20	526
5	0.1.2.1.1.2	*Bacteroidetes*[F-1]	1	69	0	66	3	0	0	0
6	0.1.2.1.1.2.1	*Bacteroidetes*[G-1]	1	69	0	66	3	0	0	0
7	0.1.2.1.1.2.1.1	sp._oral_taxon_272	0	69	0	66	3	0	0	0
5	0.1.2.1.1.3	*Bacteroidetes*[F-2]	1	3	0	1	0	2	0	0

6	0.1.2.1.1.3.1	*Bacteroidetes*[G-2]	1	3	0	1	0	2	0	0
7	0.1.2.1.1.3.1.1	sp._oral_taxon_274	0	3	0	1	0	2	0	0
5	0.1.2.1.1.4	*Bacteroidetes*[F-3]	1	8	0	0	3	5	0	0
6	0.1.2.1.1.4.1	*Bacteroidetes*[G-3]	2	8	0	0	3	5	0	0
7	0.1.2.1.1.4.1.1	sp._oral_taxon_280	0	2	0	0	2	0	0	0
7	0.1.2.1.1.4.1.3	sp._oral_taxon_365	0	6	0	0	1	5	0	0
5	0.1.2.1.1.6	*Bacteroidetes*[F-5]	1	7	0	0	0	7	0	0
6	0.1.2.1.1.6.1	*Bacteroidetes*[G-5]	1	7	0	0	0	7	0	0
7	0.1.2.1.1.6.1.3	sp._oral_taxon_511	0	7	0	0	0	7	0	0
5	0.1.2.1.1.8	*Porphyromonadaceae*	2	331	3	169	14	127	6	12
6	0.1.2.1.1.8.1	*Porphyromonas*	4	317	3	169	14	120	0	11
7	0.1.2.1.1.8.1.2	*catoniae*	0	1	0	0	0	1	0	0
7	0.1.2.1.1.8.1.3	*endodontalis*	0	112	0	40	14	58	0	0
7	0.1.2.1.1.8.1.4	*gingivalis*	0	123	3	120	0	0	0	0
7	0.1.2.1.1.8.1.13	unclassified	0	81	0	9	0	61	0	11
6	0.1.2.1.1.8.2	*Tannerella*	3	14	0	0	0	7	6	1
7	0.1.2.1.1.8.2.1	*forsythia*	0	2	0	0	0	1	0	1
7	0.1.2.1.1.8.2.3	sp._oral_taxon_286	0	6	0	0	0	0	6	0

7	0.1.2.1.1.8.2.4	unclassified	0	6	0	0	0	6	0	0
5	0.1.2.1.1.9	*Prevotellaceae*	1	4599	0	611	1871	1591	14	512
6	0.1.2.1.1.9.1	*Prevotella*	34	4599	0	611	1871	1591	14	512
7	0.1.2.1.1.9.1.1	*baroniae*	0	3	0	3	0	0	0	0
7	0.1.2.1.1.9.1.3	*buccae*	0	96	0	17	4	75	0	0
7	0.1.2.1.1.9.1.5	*dentalis*	0	10	0	8	2	0	0	0
7	0.1.2.1.1.9.1.6	*denticola*	0	1037	0	14	540	129	0	354
7	0.1.2.1.1.9.1.9	*intermedia*	0	323	0	23	300	0	0	0
7	0.1.2.1.1.9.1.11	*maculosa*	0	3	0	1	0	2	0	0
7	0.1.2.1.1.9.1.12	*marshii*	0	49	0	0	0	49	0	0
7	0.1.2.1.1.9.1.13	*melaninogenica*	0	124	0	1	0	0	14	109
7	0.1.2.1.1.9.1.14	*micans*	0	1	0	1	0	0	0	0
7	0.1.2.1.1.9.1.16	*multisaccharivorax*	0	672	0	1	671	0	0	0
7	0.1.2.1.1.9.1.17	*nigrescens*	0	406	0	81	164	159	0	2
7	0.1.2.1.1.9.1.18	*oralis*	0	197	0	101	5	76	0	15
7	0.1.2.1.1.9.1.19	*oris*	0	89	0	22	7	59	0	1
7	0.1.2.1.1.9.1.20	*oulorum*	0	19	0	1	4	13	0	1
7	0.1.2.1.1.9.1.21	*pallens*	0	5	0	1	0	3	0	1

7	0.1.2.1.1.9.1.22	*pleuritidis*_[NV]	0	8	0	8	0	0	0	0
7	0.1.2.1.1.9.1.23	*saccharolytica*	0	5	0	0	0	2	0	3
7	0.1.2.1.1.9.1.24	*salivae*	0	16	0	0	2	0	0	14
7	0.1.2.1.1.9.1.26	sp._Oral_Taxon_820	0	54	0	0	16	38	0	0
7	0.1.2.1.1.9.1.27	sp._oral_taxon_292	0	58	0	19	8	31	0	0
7	0.1.2.1.1.9.1.30	sp._oral_taxon_299	0	8	0	0	0	8	0	0
7	0.1.2.1.1.9.1.31	sp._oral_taxon_300	0	81	0	20	34	26	0	1
7	0.1.2.1.1.9.1.32	sp._oral_taxon_301	0	2	0	0	0	2	0	0
7	0.1.2.1.1.9.1.33	sp._oral_taxon_302	0	79	0	16	11	52	0	0
7	0.1.2.1.1.9.1.37	sp._oral_taxon_308	0	2	0	2	0	0	0	0
7	0.1.2.1.1.9.1.42	sp._oral_taxon_315	0	13	0	0	0	13	0	0
7	0.1.2.1.1.9.1.43	sp._oral_taxon_317	0	93	0	90	0	3	0	0
7	0.1.2.1.1.9.1.44	sp._oral_taxon_376	0	3	0	0	3	0	0	0
7	0.1.2.1.1.9.1.48	sp._oral_taxon_473	0	39	0	39	0	0	0	0
7	0.1.2.1.1.9.1.50	sp._oral_taxon_475	0	1	0	1	0	0	0	0
7	0.1.2.1.1.9.1.52	sp._oral_taxon_526	0	22	0	14	8	0	0	0
7	0.1.2.1.1.9.1.54	*tannerae*	0	1006	0	110	78	817	0	1
7	0.1.2.1.1.9.1.55	unclassified	0	73	0	17	14	32	0	10

7	0.1.2.1.1.9.1.56	*veroralis*	0	2	0	0	0	2	0	0
5	0.1.2.1.1.10	unclassified	1	12	0	7	3	0	0	2
6	0.1.2.1.1.10.1	unclassified	1	12	0	7	3	0	0	2
7	0.1.2.1.1.10.1.1	unclassified	0	12	0	7	3	0	0	2
3	0.1.2.2	*Flavobacteria*	1	181	0	81	4	47	0	49
4	0.1.2.2.1	*Flavobacteriales*	1	181	0	81	4	47	0	49
5	0.1.2.2.1.3	*Flavobacteriaceae*	2	181	0	81	4	47	0	49
6	0.1.2.2.1.3.1	*Bergeyella*	1	4	0	1	0	0	0	3
7	0.1.2.2.1.3.1.2	sp._oral_taxon_322	0	4	0	1	0	0	0	3
6	0.1.2.2.1.3.2	*Capnocytophaga*	9	177	0	80	4	47	0	46
7	0.1.2.2.1.3.2.1	*gingivalis*	0	5	0	1	3	0	0	1
7	0.1.2.2.1.3.2.2	*granulosa*	0	14	0	0	0	1	0	13
7	0.1.2.2.1.3.2.3	*haemolytica*	0	4	0	0	0	4	0	0
7	0.1.2.2.1.3.2.4	*leadbetteri*	0	73	0	47	0	10	0	16
7	0.1.2.2.1.3.2.8	sp._oral_taxon_326	0	3	0	0	0	2	0	1
7	0.1.2.2.1.3.2.9	sp._oral_taxon_332	0	2	0	1	0	1	0	0
7	0.1.2.2.1.3.2.14	sp._oral_taxon_380	0	1	0	0	0	1	0	0
7	0.1.2.2.1.3.2.16	*sputigena*	0	45	0	13	0	28	0	4

7	0.1.2.2.1.3.2.17	unclassified	0	30	0	18	1	0	0	11
2	0.1.4	*Chloroflexi*	1	1	0	1	0	0	0	0
3	0.1.4.1	*Chloroflexi*[C-1]	1	1	0	1	0	0	0	0
4	0.1.4.1.1	*Chloroflexi*[O-1]	1	1	0	1	0	0	0	0
5	0.1.4.1.1.1	*Chloroflexi*[F-1]	1	1	0	1	0	0	0	0
6	0.1.4.1.1.1.1	*Chloroflexi*[G-1]	1	1	0	1	0	0	0	0
7	0.1.4.1.1.1.1.1	sp._oral_taxon_439	0	1	0	1	0	0	0	0
2	0.1.5	*Firmicutes*	4	15684	4183	1642	1904	1264	4129	2562
3	0.1.5.1	*Bacilli*	2	11357	4180	933	144	127	3789	2184
4	0.1.5.1.1	*Bacillales*	1	88	7	60	0	11	0	10
5	0.1.5.1.1.4	*Staphylococcaceae*	1	88	7	60	0	11	0	10
6	0.1.5.1.1.4.1	*Gemella*	4	88	7	60	0	11	0	10
7	0.1.5.1.1.4.1.2	*haemolysans*	0	4	4	0	0	0	0	0
7	0.1.5.1.1.4.1.3	*morbillorum*	0	76	1	58	0	7	0	10
7	0.1.5.1.1.4.1.4	*sanguinis*	0	4	0	2	0	2	0	0
7	0.1.5.1.1.4.1.5	unclassified	0	4	2	0	0	2	0	0
4	0.1.5.1.2	*Lactobacillales*	5	11269	4173	873	144	116	3789	2174
5	0.1.5.1.2.1	*Aerococcaceae*	1	2	0	0	0	2	0	0

6	0.1.5.1.2.1.1	*Abiotrophia*	1	2	0	0	0	2	0	0
7	0.1.5.1.2.1.1.1	*defectiva*	0	2	0	0	0	2	0	0
5	0.1.5.1.2.2	*Carnobacteriaceae*	1	46	0	14	3	11	2	16
6	0.1.5.1.2.2.3	*Granulicatella*	1	46	0	14	3	11	2	16
7	0.1.5.1.2.2.3.1	*adiacens*	0	46	0	14	3	11	2	16
5	0.1.5.1.2.3	*Enterococcaceae*	1	3	3	0	0	0	0	0
6	0.1.5.1.2.3.1	*Enterococcus*	1	3	3	0	0	0	0	0
7	0.1.5.1.2.3.1.2	*faecalis*	0	3	3	0	0	0	0	0
5	0.1.5.1.2.4	*Lactobacillaceae*	1	7448	4063	10	11	16	3111	237
6	0.1.5.1.2.4.1	*Lactobacillus*	11	7448	4063	10	11	16	3111	237
7	0.1.5.1.2.4.1.3	*casei*	0	222	68	0	0	0	154	0
7	0.1.5.1.2.4.1.5	*crispatus*	0	742	741	0	0	0	1	0
7	0.1.5.1.2.4.1.6	*fermentum*	0	215	16	0	0	0	199	0
7	0.1.5.1.2.4.1.7	*gasseri*	0	5108	2490	1	4	12	2506	95
7	0.1.5.1.2.4.1.11	*kisonensis*	0	1	1	0	0	0	0	0
7	0.1.5.1.2.4.1.12	*oris*	0	23	0	0	0	0	23	0
7	0.1.5.1.2.4.1.13	*paracasei*	0	36	0	0	0	0	36	0
7	0.1.5.1.2.4.1.14	*parafarraginis*	0	180	177	0	0	3	0	0

7	0.1.5.1.2.4.1.17	*salivarius*	0	8	0	1	7	0	0	0
7	0.1.5.1.2.4.1.20	unclassified	0	328	105	8	0	1	76	138
7	0.1.5.1.2.4.1.21	*vaginalis*	0	585	465	0	0	0	116	4
5	0.1.5.1.2.5	*Streptococcaceae*	2	3770	107	849	130	87	676	1921
6	0.1.5.1.2.5.1	*Lactococcus*	1	11	11	0	0	0	0	0
7	0.1.5.1.2.5.1.1	*lactis*	0	11	11	0	0	0	0	0
6	0.1.5.1.2.5.2	*Streptococcus*	16	3759	96	849	130	87	676	1921
7	0.1.5.1.2.5.2.2	*anginosus*	0	86	0	61	16	4	0	5
7	0.1.5.1.2.5.2.6	*gordonii*	0	265	4	53	54	4	2	148
7	0.1.5.1.2.5.2.9	*mitis*	0	8	1	0	0	7	0	0
7	0.1.5.1.2.5.2.10	*mitis*_bv_2	0	48	2	3	0	25	2	16
7	0.1.5.1.2.5.2.11	*mutans*	0	1571	0	45	4	15	607	900
7	0.1.5.1.2.5.2.14	*parasanguinis*_II	0	2	0	2	0	0	0	0
7	0.1.5.1.2.5.2.25	sp._oral_taxon_058	0	26	0	23	0	3	0	0
7	0.1.5.1.2.5.2.26	sp._oral_taxon_061	0	1	0	0	0	1	0	0
7	0.1.5.1.2.5.2.27	sp._oral_taxon_064	0	59	0	46	6	5	2	0
7	0.1.5.1.2.5.2.28	sp._oral_taxon_065	0	14	0	12	0	0	2	0
7	0.1.5.1.2.5.2.33	sp._oral_taxon_070	0	4	0	0	0	0	0	4

7	0.1.5.1.2.5.2.34	sp._oral_taxon_071	0	28	1	1	7	2	14	3
7	0.1.5.1.2.5.2.35	sp._oral_taxon_073	0	4	0	3	0	0	0	1
7	0.1.5.1.2.5.2.36	sp._oral_taxon_074	0	5	0	0	0	5	0	0
7	0.1.5.1.2.5.2.41	unclassified	0	1637	87	600	43	16	47	844
7	0.1.5.1.2.5.2.42	*vestibularis*	0	1	1	0	0	0	0	0
3	0.1.5.2	*Clostridia*	1	4204	3	659	1728	1096	340	378
4	0.1.5.2.1	*Clostridiales*	8	4204	3	659	1728	1096	340	378
5	0.1.5.2.1.1	*Clostridiales*[F-1]	1	3	0	1	0	2	0	0
6	0.1.5.2.1.1.1	*Clostridiales*[F-1][G-1]	1	3	0	1	0	2	0	0
7	0.1.5.2.1.1.1.1	sp._oral_taxon_093	0	3	0	1	0	2	0	0
5	0.1.5.2.1.2	*Clostridiales*[F-2]	1	16	0	1	0	15	0	0
6	0.1.5.2.1.2.1	*Clostridiales*[F-2][G-1]	1	16	0	1	0	15	0	0
7	0.1.5.2.1.2.1.1	sp._oral_taxon_075	0	16	0	1	0	15	0	0
5	0.1.5.2.1.3	*Eubacteriaceae*[15]	1	1118	0	108	1009	1	0	0
6	0.1.5.2.1.3.2	*Pseudoramibacter*	1	1118	0	108	1009	1	0	0
7	0.1.5.2.1.3.2.1	*alactolyticus*	0	1118	0	108	1009	1	0	0
5	0.1.5.2.1.4	*Lachnospiraceae*[14]	11	296	0	73	162	44	0	17
6	0.1.5.2.1.4.2	*Catonella*	3	83	0	61	15	7	0	0

7	0.1.5.2.1.4.2.1	*morbi*	0	68	0	48	15	5	0	0
7	0.1.5.2.1.4.2.2	sp._oral_taxon_164	0	2	0	0	0	2	0	0
7	0.1.5.2.1.4.2.3	sp._oral_taxon_451	0	13	0	13	0	0	0	0
6	0.1.5.2.1.4.3	*Eubacterium*[14][G-1]	1	7	0	1	0	2	0	4
7	0.1.5.2.1.4.3.1	*saburreum*	0	7	0	1	0	2	0	4
6	0.1.5.2.1.4.4	*Johnsonella*	1	2	0	0	0	2	0	0
7	0.1.5.2.1.4.4.2	sp._oral_taxon_166	0	2	0	0	0	2	0	0
6	0.1.5.2.1.4.5	*Lachnospiraceae*[G-1]	3	12	0	0	0	9	0	3
7	0.1.5.2.1.4.5.2	sp._oral_taxon_083	0	5	0	0	0	5	0	0
7	0.1.5.2.1.4.5.3	sp._oral_taxon_089	0	4	0	0	0	4	0	0
7	0.1.5.2.1.4.5.4	sp._oral_taxon_107	0	3	0	0	0	0	0	3
6	0.1.5.2.1.4.6	*Lachnospiraceae*[G-2]	1	3	0	2	0	1	0	0
7	0.1.5.2.1.4.6.2	sp._oral_taxon_100	0	3	0	2	0	1	0	0
6	0.1.5.2.1.4.8	*Lachnospiraceae*[G-4]	1	19	0	0	16	0	0	3
7	0.1.5.2.1.4.8.4	unclassified	0	19	0	0	16	0	0	3
6	0.1.5.2.1.4.11	*Lachnospiraceae*[G-7]	1	1	0	1	0	0	0	0
7	0.1.5.2.1.4.11.2	sp._oral_taxon_163	0	1	0	1	0	0	0	0
6	0.1.5.2.1.4.12	*Lachnospiraceae*[G-8]	1	2	0	2	0	0	0	0

7	0.1.5.2.1.4.12.1	sp._oral_taxon_500	0	2	0	2	0	0	0	0
6	0.1.5.2.1.4.13	*Oribacterium*	1	2	0	0	0	2	0	0
7	0.1.5.2.1.4.13.6	unclassified	0	2	0	0	0	2	0	0
6	0.1.5.2.1.4.14	*Shuttleworthia*	1	155	0	3	131	20	0	1
7	0.1.5.2.1.4.14.1	*satelles*	0	155	0	3	131	20	0	1
6	0.1.5.2.1.4.15	unclassified	1	10	0	3	0	1	0	6
7	0.1.5.2.1.4.15.1	unclassified	0	10	0	3	0	1	0	6
5	0.1.5.2.1.5	*Peptococcaceae*	1	3	0	2	1	0	0	0
6	0.1.5.2.1.5.1	*Peptococcus*	2	3	0	2	1	0	0	0
7	0.1.5.2.1.5.1.1	sp._oral_taxon_167	0	2	0	1	1	0	0	0
7	0.1.5.2.1.5.1.3	unclassified	0	1	0	1	0	0	0	0
5	0.1.5.2.1.6	*Peptostreptococcaceae*[11]	12	936	0	125	228	582	0	1
6	0.1.5.2.1.6.1	*Eubacterium*[11][G-1]	2	424	0	11	87	326	0	0
7	0.1.5.2.1.6.1.1	*infirmum*	0	421	0	8	87	326	0	0
7	0.1.5.2.1.6.1.2	*sulci*	0	3	0	3	0	0	0	0
6	0.1.5.2.1.6.2	*Eubacterium*[11][G-3]	1	17	0	0	0	17	0	0
7	0.1.5.2.1.6.2.1	*brachy*	0	17	0	0	0	17	0	0
6	0.1.5.2.1.6.3	*Eubacterium*[11][G-5]	1	26	0	0	1	25	0	0

7	0.1.5.2.1.6.3.1	*saphenum*	0	26	0	0	1	25	0	0
6	0.1.5.2.1.6.4	*Eubacterium*[11][G-6]	1	3	0	3	0	0	0	0
7	0.1.5.2.1.6.4.1	*minutum*	0	3	0	3	0	0	0	0
6	0.1.5.2.1.6.5	*Eubacterium*[11][G-7]	1	30	0	30	0	0	0	0
7	0.1.5.2.1.6.5.2	*yurii*	0	30	0	30	0	0	0	0
6	0.1.5.2.1.6.6	*Filifactor*	1	123	0	20	23	80	0	0
7	0.1.5.2.1.6.6.1	*alocis*	0	123	0	20	23	80	0	0
6	0.1.5.2.1.6.7	*Mogibacterium*	2	146	0	53	30	62	0	1
7	0.1.5.2.1.6.7.4	*timidum*	0	55	0	8	30	17	0	0
7	0.1.5.2.1.6.7.5	unclassified	0	91	0	45	0	45	0	1
6	0.1.5.2.1.6.9	*Peptostreptococcaceae*[11][G-2]	1	10	0	1	5	4	0	0
7	0.1.5.2.1.6.9.1	sp._oral_taxon_091	0	10	0	1	5	4	0	0
6	0.1.5.2.1.6.11	*Peptostreptococcaceae*[11][G-4]	2	7	0	1	0	6	0	0
7	0.1.5.2.1.6.11.1	sp._oral_taxon_103	0	3	0	1	0	2	0	0
7	0.1.5.2.1.6.11.3	unclassified	0	4	0	0	0	4	0	0
6	0.1.5.2.1.6.12	*Peptostreptococcaceae*[11][G-5]	1	27	0	0	0	27	0	0
7	0.1.5.2.1.6.12.1	sp._oral_taxon_493	0	27	0	0	0	27	0	0
6	0.1.5.2.1.6.13	*Peptostreptococcaceae*[11][G-7]	1	21	0	0	0	21	0	0

7	0.1.5.2.1.6.13.1	sp._oral_taxon_081	0	21	0	0	0	21	0	0
6	0.1.5.2.1.6.14	*Peptostreptococcus*	1	102	0	6	82	14	0	0
7	0.1.5.2.1.6.14.2	*stomatis*	0	102	0	6	82	14	0	0
5	0.1.5.2.1.7	*Peptostreptococcaceae*[13]	1	39	0	7	22	10	0	0
6	0.1.5.2.1.7.3	*Parvimonas*	2	39	0	7	22	10	0	0
7	0.1.5.2.1.7.3.1	*micra*	0	36	0	4	22	10	0	0
7	0.1.5.2.1.7.3.4	unclassified	0	3	0	3	0	0	0	0
5	0.1.5.2.1.9	*Veillonellaceae*	7	1793	3	342	306	442	340	360
6	0.1.5.2.1.9.1	*Anaeroglobus*	1	32	0	32	0	0	0	0
7	0.1.5.2.1.9.1.1	*geminatus*	0	32	0	32	0	0	0	0
6	0.1.5.2.1.9.3	*Dialister*	2	445	0	37	219	177	0	12
7	0.1.5.2.1.9.3.1	*invisus*	0	344	0	34	147	151	0	12
7	0.1.5.2.1.9.3.3	*pneumosintes*	0	101	0	3	72	26	0	0
6	0.1.5.2.1.9.4	*Megasphaera*	2	78	0	4	0	24	0	50
7	0.1.5.2.1.9.4.1	*micronuciformis*	0	61	0	4	0	7	0	50
7	0.1.5.2.1.9.4.2	sp._oral_taxon_123	0	17	0	0	0	17	0	0
6	0.1.5.2.1.9.6	*Selenomonas*	7	413	1	192	13	162	2	43
7	0.1.5.2.1.9.6.1	*artemidis*_oral_taxon_124	0	24	0	24	0	0	0	0

7	0.1.5.2.1.9.6.2	*dianae*	0	96	0	57	6	27	1	5
7	0.1.5.2.1.9.6.3	*flueggei*	0	10	0	0	0	10	0	0
7	0.1.5.2.1.9.6.5	*noxia*	0	47	0	15	0	8	1	23
7	0.1.5.2.1.9.6.6	sp._oral_taxon_126	0	2	0	2	0	0	0	0
7	0.1.5.2.1.9.6.22	*sputigena*	0	137	0	47	0	84	0	6
7	0.1.5.2.1.9.6.23	unclassified	0	97	1	47	7	33	0	9
6	0.1.5.2.1.9.7	*Veillonella*	2	691	2	42	44	11	338	254
7	0.1.5.2.1.9.7.3	*parvula*	0	650	2	31	42	8	314	253
7	0.1.5.2.1.9.7.6	unclassified	0	41	0	11	2	3	24	1
6	0.1.5.2.1.9.8	*Veillonellaceae*[G-1]	5	121	0	30	30	60	0	1
7	0.1.5.2.1.9.8.1	sp._oral_taxon_129	0	12	0	0	8	4	0	0
7	0.1.5.2.1.9.8.2	sp._oral_taxon_132	0	20	0	6	2	12	0	0
7	0.1.5.2.1.9.8.6	sp._oral_taxon_150	0	20	0	13	7	0	0	0
7	0.1.5.2.1.9.8.7	sp._oral_taxon_155	0	55	0	8	2	44	0	1
7	0.1.5.2.1.9.8.9	unclassified	0	14	0	3	11	0	0	0
6	0.1.5.2.1.9.9	unclassified	1	13	0	5	0	8	0	0
7	0.1.5.2.1.9.9.1	unclassified	0	13	0	5	0	8	0	0
3	0.1.5.3	*Mollicutes*	2	63	0	34	26	3	0	0

4	0.1.5.3.2	*Anaeroplasmatales*	2	62	0	34	25	3	0	0
5	0.1.5.3.2.1	*Anaeroplasmataceae*	1	1	0	0	0	1	0	0
6	0.1.5.3.2.1.1	*Lactobacillus*[17]	1	1	0	0	0	1	0	0
7	0.1.5.3.2.1.1.1	*catenaformis*	0	1	0	0	0	1	0	0
5	0.1.5.3.2.2	*Erysipelotrichaceae*	2	61	0	34	25	2	0	0
6	0.1.5.3.2.2.1	*Bulleidia*	1	1	0	0	1	0	0	0
7	0.1.5.3.2.2.1.1	*extructa*	0	1	0	0	1	0	0	0
6	0.1.5.3.2.2.3	*Solobacterium*	1	60	0	34	24	2	0	0
7	0.1.5.3.2.2.3.1	*moorei*	0	60	0	34	24	2	0	0
4	0.1.5.3.3	*Mycoplasmatales*	1	1	0	0	1	0	0	0
5	0.1.5.3.3.1	*Mycoplasmataceae*	1	1	0	0	1	0	0	0
6	0.1.5.3.3.1.1	*Mycoplasma*	1	1	0	0	1	0	0	0
7	0.1.5.3.3.1.1.2	*faucium*	0	1	0	0	1	0	0	0
3	0.1.5.4	unclassified	1	60	0	16	6	38	0	0
4	0.1.5.4.1	unclassified	1	60	0	16	6	38	0	0
5	0.1.5.4.1.1	unclassified	1	60	0	16	6	38	0	0
6	0.1.5.4.1.1.1	unclassified	1	60	0	16	6	38	0	0
7	0.1.5.4.1.1.1.1	unclassified	0	60	0	16	6	38	0	0

2	0.1.6	*Fusobacteria*	1	1245	6	375	99	453	0	312
3	0.1.6.1	*Fusobacteria*	1	1245	6	375	99	453	0	312
4	0.1.6.1.1	*Fusobacteriales*	1	1245	6	375	99	453	0	312
5	0.1.6.1.1.1	*Fusobacteriaceae*	4	1245	6	375	99	453	0	312
6	0.1.6.1.1.1.1	*Fusobacterium*	7	1097	1	363	93	402	0	238
7	0.1.6.1.1.1.1.2	*nucleatum*_subsp._*animalis*	0	527	0	112	82	293	0	40
7	0.1.6.1.1.1.1.4	*nucleatum*_subsp._*polymorphum*	0	337	0	102	0	47	0	188
7	0.1.6.1.1.1.1.5	*nucleatum*_subsp._*vincentii*	0	117	1	107	9	0	0	0
7	0.1.6.1.1.1.1.6	*periodonticum*	0	57	0	24	0	33	0	0
7	0.1.6.1.1.1.1.7	sp._oral_taxon_203	0	1	0	1	0	0	0	0
7	0.1.6.1.1.1.1.9	sp._oral_taxon_370	0	1	0	0	1	0	0	0
7	0.1.6.1.1.1.1.10	unclassified	0	57	0	17	1	29	0	10
6	0.1.6.1.1.1.2	*Leptotrichia*	12	144	5	12	2	51	0	74
7	0.1.6.1.1.1.2.1	*buccalis*	0	5	0	1	0	3	0	1
7	0.1.6.1.1.1.2.3	*hofstadii*	0	1	0	0	1	0	0	0
7	0.1.6.1.1.1.2.4	*hongkongensis*	0	2	0	0	0	0	0	2
7	0.1.6.1.1.1.2.5	*shahii*	0	7	0	7	0	0	0	0
7	0.1.6.1.1.1.2.7	sp._oral_taxon_212	0	22	0	0	0	0	0	22

7	0.1.6.1.1.1.2.8	sp._oral_taxon_215	0	22	0	1	0	0	0	21
7	0.1.6.1.1.1.2.14	sp._oral_taxon_223	0	3	1	0	0	0	0	2
7	0.1.6.1.1.1.2.15	sp._oral_taxon_225	0	1	0	0	0	0	0	1
7	0.1.6.1.1.1.2.16	sp._oral_taxon_392	0	4	0	0	0	0	0	4
7	0.1.6.1.1.1.2.19	sp._oral_taxon_463	0	8	0	0	0	8	0	0
7	0.1.6.1.1.1.2.20	sp._oral_taxon_498	0	34	0	1	0	27	0	6
7	0.1.6.1.1.1.2.21	unclassified	0	35	4	2	1	13	0	15
6	0.1.6.1.1.1.3	*Sneathia*	1	2	0	0	2	0	0	0
7	0.1.6.1.1.1.3.2	*sanguinegens*	0	2	0	0	2	0	0	0
6	0.1.6.1.1.1.4	unclassified	1	2	0	0	2	0	0	0
7	0.1.6.1.1.1.4.1	unclassified	0	2	0	0	2	0	0	0
2	0.1.7	*Proteobacteria*	4	1598	38	888	60	162	43	407
3	0.1.7.2	*Betaproteobacteria*	2	982	6	580	51	83	7	255
4	0.1.7.2.1	*Burkholderiales*	2	11	0	6	0	0	0	5
5	0.1.7.2.1.2	*Burkholderiaceae*	1	6	0	1	0	0	0	5
6	0.1.7.2.1.2.2	*Lautropia*	1	6	0	1	0	0	0	5
7	0.1.7.2.1.2.2.1	*mirabilis*	0	6	0	1	0	0	0	5
5	0.1.7.2.1.3	*Comomonadaceae*	1	5	0	5	0	0	0	0

6	0.1.7.2.1.3.4	unclassified	1	5	0	5	0	0	0	0
7	0.1.7.2.1.3.4.1	unclassified	0	5	0	5	0	0	0	0
4	0.1.7.2.2	*Neisseriales*	1	971	6	574	51	83	7	250
5	0.1.7.2.2.1	*Neisseriaceae*	4	971	6	574	51	83	7	250
6	0.1.7.2.2.1.1	*Eikenella*	1	54	0	24	0	6	0	24
7	0.1.7.2.2.1.1.1	*corrodens*	0	54	0	24	0	6	0	24
6	0.1.7.2.2.1.2	*Kingella*	2	69	0	2	0	16	5	46
7	0.1.7.2.2.1.2.1	*denitrificans*	0	21	0	2	0	16	0	3
7	0.1.7.2.2.1.2.3	*oralis*	0	48	0	0	0	0	5	43
6	0.1.7.2.2.1.3	*Neisseria*	6	847	5	548	51	61	2	180
7	0.1.7.2.2.1.3.1	*bacilliformis*	0	97	1	1	0	0	0	95
7	0.1.7.2.2.1.3.2	*elongata*	0	113	3	68	13	0	0	29
7	0.1.7.2.2.1.3.3	*flava*	0	64	0	0	0	18	1	45
7	0.1.7.2.2.1.3.12	sp._oral_taxon_014	0	27	0	0	0	26	0	1
7	0.1.7.2.2.1.3.15	sp._oral_taxon_018	0	1	0	0	0	0	1	0
7	0.1.7.2.2.1.3.20	unclassified	0	545	1	479	38	17	0	10
6	0.1.7.2.2.1.4	*Simonsiella*	1	1	1	0	0	0	0	0
7	0.1.7.2.2.1.4.1	*muelleri*	0	1	1	0	0	0	0	0

3	0.1.7.3	*Deltaproteobacteria*	1	2	0	2	0	0	0	0
4	0.1.7.3.2	*Desulfobacterales*	1	2	0	2	0	0	0	0
5	0.1.7.3.2.1	*Desulfobulbaceae*	1	2	0	2	0	0	0	0
6	0.1.7.3.2.1.1	*Desulfobulbus*	1	2	0	2	0	0	0	0
7	0.1.7.3.2.1.1.1	sp._oral_taxon_041	0	2	0	2	0	0	0	0
3	0.1.7.4	*Epsilonproteobacteria*	1	248	4	92	5	36	3	108
4	0.1.7.4.1	*Campylobacterales*	1	248	4	92	5	36	3	108
5	0.1.7.4.1.1	*Campylobacteraceae*	1	248	4	92	5	36	3	108
6	0.1.7.4.1.1.2	*Campylobacter*	3	248	4	92	5	36	3	108
7	0.1.7.4.1.1.2.1	*concisus*	0	38	0	33	0	0	3	2
7	0.1.7.4.1.1.2.3	*gracilis*	0	112	1	38	2	12	0	59
7	0.1.7.4.1.1.2.5	*showae*	0	98	3	21	3	24	0	47
3	0.1.7.5	*Gammaproteobacteria*	4	366	28	214	4	43	33	44
4	0.1.7.5.1	*Cardiobacteriales*	1	9	0	2	1	1	0	5
5	0.1.7.5.1.1	*Cardiobacteriaceae*	1	9	0	2	1	1	0	5
6	0.1.7.5.1.1.1	*Cardiobacterium*	2	9	0	2	1	1	0	5
7	0.1.7.5.1.1.1.1	*hominis*	0	6	0	0	1	1	0	4
7	0.1.7.5.1.1.1.3	*valvarum*	0	3	0	2	0	0	0	1

4	0.1.7.5.2	*Enterobacteriales*	1	18	14	1	1	1	0	1
5	0.1.7.5.2.1	*Enterobacteriaceae*	1	18	14	1	1	1	0	1
6	0.1.7.5.2.1.4	*Proteus*	1	18	14	1	1	1	0	1
7	0.1.7.5.2.1.4.1	*mirabilis*	0	18	14	1	1	1	0	1
4	0.1.7.5.3	*Pasteurellales*	1	338	13	211	2	41	33	38
5	0.1.7.5.3.1	*Pasteurellaceae*	3	338	13	211	2	41	33	38
6	0.1.7.5.3.1.1	*Aggregatibacter*	2	219	0	151	0	40	6	22
7	0.1.7.5.3.1.1.5	sp._oral_taxon_458	0	2	0	1	0	0	0	1
7	0.1.7.5.3.1.1.8	unclassified	0	217	0	150	0	40	6	21
6	0.1.7.5.3.1.2	*Haemophilus*	3	46	13	27	0	1	4	1
7	0.1.7.5.3.1.2.4	*parainfluenzae*	0	27	0	22	0	0	4	1
7	0.1.7.5.3.1.2.6	sp._oral_taxon_036	0	18	13	5	0	0	0	0
7	0.1.7.5.3.1.2.7	unclassified	0	1	0	0	0	1	0	0
6	0.1.7.5.3.1.4	unclassified	1	73	0	33	2	0	23	15
7	0.1.7.5.3.1.4.1	unclassified	0	73	0	33	2	0	23	15
4	0.1.7.5.4	*Pseudomonadales*	1	1	1	0	0	0	0	0
5	0.1.7.5.4.2	*Pseudomonadaceae*	1	1	1	0	0	0	0	0
6	0.1.7.5.4.2.1	*Pseudomonas*	1	1	1	0	0	0	0	0

7	0.1.7.5.4.2.1.2	*fluorescens*	0	1	1	0	0	0	0	0
2	0.1.8	SR1	1	2	0	0	0	0	0	2
3	0.1.8.1	SR1[C-1]	1	2	0	0	0	0	0	2
4	0.1.8.1.1	SR1[O-1]	1	2	0	0	0	0	0	2
5	0.1.8.1.1.1	SR1[F-1]	1	2	0	0	0	0	0	2
6	0.1.8.1.1.1.1	SR1[G-1]	1	2	0	0	0	0	0	2
7	0.1.8.1.1.1.1.1	sp._oral_taxon_345	0	2	0	0	0	0	0	2
2	0.1.9	*Spirochaetes*	1	170	0	38	19	113	0	0
3	0.1.9.1	*Spirochaetes*	1	170	0	38	19	113	0	0
4	0.1.9.1.1	*Spirochaetales*	1	170	0	38	19	113	0	0
5	0.1.9.1.1.1	*Spirochaetaceae*	1	170	0	38	19	113	0	0
6	0.1.9.1.1.1.1	*Treponema*	12	170	0	38	19	113	0	0
7	0.1.9.1.1.1.1.1	*amylovorum*	0	49	0	3	0	46	0	0
7	0.1.9.1.1.1.1.2	*denticola*	0	2	0	0	0	2	0	0
7	0.1.9.1.1.1.1.4	*maltophilum*	0	9	0	0	0	9	0	0
7	0.1.9.1.1.1.1.7	*parvum*	0	1	0	0	0	1	0	0
7	0.1.9.1.1.1.1.8	*pectinovorum*	0	2	0	0	1	1	0	0
7	0.1.9.1.1.1.1.10	*socranskii*	0	77	0	24	3	50	0	0

7	0.1.9.1.1.1.1.11	sp._oral_taxon_226	0	1	0	1	0	0	0	0
7	0.1.9.1.1.1.1.21	sp._oral_taxon_238	0	5	0	0	1	4	0	0
7	0.1.9.1.1.1.1.34	sp._oral_taxon_257	0	14	0	0	14	0	0	0
7	0.1.9.1.1.1.1.35	sp._oral_taxon_258	0	1	0	1	0	0	0	0
7	0.1.9.1.1.1.1.41	sp._oral_taxon_268	0	6	0	6	0	0	0	0
7	0.1.9.1.1.1.1.49	unclassified	0	3	0	3	0	0	0	0
2	0.1.10	*Synergistetes*	1	75	0	13	3	59	0	0
3	0.1.10.1	*Synergistetes*[C-1]	1	75	0	13	3	59	0	0
4	0.1.10.1.1	*Synergistetes*[O-1]	1	75	0	13	3	59	0	0
5	0.1.10.1.1.2	*Synergistetes*[F-2]	1	75	0	13	3	59	0	0
6	0.1.10.1.1.2.1	*Synergistetes*[G-3]	6	75	0	13	3	59	0	0
7	0.1.10.1.1.2.1.2	sp._oral_taxon_359	0	6	0	0	0	6	0	0
7	0.1.10.1.1.2.1.3	sp._oral_taxon_360	0	3	0	0	0	3	0	0
7	0.1.10.1.1.2.1.4	sp._oral_taxon_361	0	1	0	0	0	1	0	0
7	0.1.10.1.1.2.1.5	sp._oral_taxon_362	0	1	0	0	0	1	0	0
7	0.1.10.1.1.2.1.6	sp._oral_taxon_363	0	52	0	6	3	43	0	0
7	0.1.10.1.1.2.1.9	unclassified	0	12	0	7	0	5	0	0
2	0.1.11	TM7	1	134	1	83	0	46	0	4

3	0.1.11.1	TM7[C-1]	1	134	1	83	0	46	0	4
4	0.1.11.1.1	TM7[O-1]	2	134	1	83	0	46	0	4
5	0.1.11.1.1.1	TM7[F-1]	3	41	1	29	0	7	0	4
6	0.1.11.1.1.1.1	TM7[G-1]	5	37	1	27	0	5	0	4
7	0.1.11.1.1.1.1.1	sp._oral_taxon_346	0	9	0	6	0	0	0	3
7	0.1.11.1.1.1.1.3	sp._oral_taxon_348	0	5	0	0	0	5	0	0
7	0.1.11.1.1.1.1.4	sp._oral_taxon_349	0	10	1	9	0	0	0	0
7	0.1.11.1.1.1.1.5	sp._oral_taxon_352	0	1	0	0	0	0	0	1
7	0.1.11.1.1.1.1.8	unclassified	0	12	0	12	0	0	0	0
6	0.1.11.1.1.1.3	TM7[G-3]	1	2	0	0	0	2	0	0
7	0.1.11.1.1.1.3.1	sp._oral_taxon_351	0	2	0	0	0	2	0	0
6	0.1.11.1.1.1.4	TM7[G-4]	1	2	0	2	0	0	0	0
7	0.1.11.1.1.1.4.1	sp._oral_taxon_355	0	2	0	2	0	0	0	0
5	0.1.11.1.1.2	TM7[F-2]	1	93	0	54	0	39	0	0
6	0.1.11.1.1.2.1	TM7[G-5]	1	93	0	54	0	39	0	0
7	0.1.11.1.1.2.1.1	sp._oral_taxon_356	0	93	0	54	0	39	0	0
2	0.1.12	unclassified	1	1	1	0	0	0	0	0
3	0.1.12.1	unclassified	1	1	1	0	0	0	0	0

4	0.1.12.1.1	unclassified	1	1	1	0	0	0	0	0
5	0.1.12.1.1.1	unclassified	1	1	1	0	0	0	0	0
6	0.1.12.1.1.1.1	unclassified	1	1	1	0	0	0	0	0
7	0.1.12.1.1.1.1.1	unclassified	0	1	1	0	0	0	0	0